

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 23, 2003, 22:21:42 : Search time 415 Seconds

(without alignments)
4851.646 Million cell updates/sec

Title: US-09-874-198-2

Perfect score: 3758

Sequence: 1 MRLTLGLGCGSVATPLGP.....GYTKVINYIPMIENITSDP 686

Scoring table:

BLOSUM62	
Xgapop 10.0, Ygapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Delop 6.0, Delext 7.0	
Delop 6.0, Delext 7.0	

Described: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 75 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=alp
-Q/cg2_1/USFTO.spool/US09874198/runat_22012003_085000_10851/app_query.fasta_1.839
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=75
-DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USFR=US09874198.ecgn.1.1.3637 @runat_22012003_085000_10851 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-MARKTIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: GenEmbl:*
2: gb_ba:*
3: gb_htg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sts:*
13: gb_sy:*
14: gb_un:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3745	99.7	2450	9 HSMASp2	X98400 H. sapiens m
2	3745	99.7	2455	9 HSMASp2R	X09926 Homo sapien
3	3298.5	87.8	4464	9 HOSAI8285	Y18286 Homo sapien
4	3285.5	87.4	4465	9 HOSAI8287	Y18287 Homo sapien
5	3065.5	81.6	3070	10 BC013893	BC013893 Mus muscu
6	3062.5	81.5	2090	10 AB009459	AB009459 Mus muscu
7	3054.5	81.3	2037	10 RNO277747	AJ277747 Rattus nor
8	2901.5	77.2	1937	10 RNO18572	Y18572 Rattus norv
9	2802.5	74.6	1898	10 RNO18564	Y18564 Rattus norv
10	2427.5	64.6	1657	10 RNO18566	Y18566 Rattus norv
11	1815.5	48.3	2282	5 AB009072	AB009072 Xenopus l
12	1634.5	43.5	1112	10 RNO18571	Y18571 Rattus norv
13	1626	43.3	1145	10 MMU19163	AF284421 Mus musculi
14	1609	42.8	3895	9 AF284421	AF284421 Homo sapi
15	1608	42.8	3863	6 AX254423	AX254423 Sequence
16	1603.5	42.7	3736	10 AB049755	AB049755 Mus muscu
17	1600	42.6	3895	6 AX156466	AX156466 Sequence
18	1555	41.4	2108	10 RNO277423	AJ277423 Rattus no
19	1553	41.3	4489	9 D17525	D17525 Human mRNA
20	1549.5	41.2	5185	10 MUSCRARF	D16492 Mouse mRNA
21	1540	41.0	2787	9 HUMMASP	D28593 Human mRNA
22	1513.5	40.3	1027	10 RNO18565	Y18565 Rattus norv
23	1510	40.2	2407	5 AB009074	AB009074 Trilakis s
24	1492	39.7	2550	5 D83276	D83276 Xenopus lae
25	1462	38.9	3229	5 AB009073	AB009073 Cyprinus
26	1435	38.2	3708	5 AB009075	AB009075 Lampetra
27	1412	37.6	2492	6 AR130745	AR130745 Sequence
28	1412	37.6	2493	6 AX302547	AX302547 Sequence
29	1412	37.6	2493	6 AX409609	AX409609 Sequence
30	1412	37.6	2493	6 HUMC1R	M14058 Human compl
31	1410	37.5	2386	6 AX250080	AX250080 Sequence
32	1410	37.5	2386	9 HSC1R	X04701 Human mRNA
33	1384	36.8	2416	10 AF148216	AF148216 Mus muscu
34	1381	36.7	2409	10 BC004637	BC004637 Mus muscu
35	1376	36.6	112769	9 HSCJ35E18	AL109811 Human DNA
36	1376	36.6	148049	9 HSAJ300188	AJ300188 Homo sapi
37	1360	36.2	1386	9 AF321562	AF321562 Homo sapi
38	1346	35.8	3519	9 AK024951	AK024951 Homo sapi
39	1281	34.1	224312	2 AL606969	AL606969 Mus muscu
40	1244.5	33.1	2711	10 MACASP	X16160 Mesocricetu
41	1238	32.9	2582	9 HUMC1SAB	M18767 Human compl
42	1238	32.9	2647	6 AX409458	AJ409458 Sequence
43	1238	32.9	2647	9 HUMC1RS	J04080 Human compl
44	1233.5	32.8	205573	10 AL591032	AL591032 Mouse DNA
45	1233	32.8	3073	9 AK055183	AK055183 Homo sapi
46	1232	32.8	2855	10 BC018319	BC018319 Mus muscu
47	1232	32.8	2915	10 BC022123	BC022123 Mus muscu
48	1230	32.7	2616	9 HSC1S	X06596 Human mRNA
49	1217	32.4	2915	10 BC027183	BC027183 Mus muscu
50	1214	32.3	2908	6 AX401672	AX401672 Sequence

QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValIleThrPheValThr 280
 DB 817 AAGACATTCCTTACTGACGATTGTAACAAAAGACACGGTGACCTTCCTTGGTTCACA 876
 QY 281 AspGluSerGlyAspHisThrGlyTrpLysIleHisTyrThrSerThrIleGlnProCys 300
 DB 877 GATGATCAGGAGACACACAGGGCTGGAAGATCCACTACACGACGACGACGCGCTTGC 936
 QY 301 ProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIleLeu 320
 DB 937 CCTATCCCATGCGCCACCTAATGGCCACGTTTCACCTGTGCACCAATAATACATCTG 996
 QY 321 LysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuGlnGlyIleLeuPro 340
 DB 997 AAGACAGCTTCTCCATCTTTGGCAGACTGGCTATGACCTTCTGCAAGGTCACTTGGCC 1056
 QY 341 LeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspArgProMetProAla 360
 DB 057 CTGAATTCCTTACTGACGATTGTGCAAAAGATGATCTTGGACCGGCAATGGCCGG 1116
 QY 361 CysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluTyrIle 380
 DB 1117 TGCAGCATTTGTGACTGTGCGCTCCTGATGATCTACCCAGTGGCCGAGTGAATACATC 1176
 QY 381 ThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPhe 400
 DB 1177 ACAGGCTCCGAGGTGACCCCTACAAAGCTGTGATTCAGTACGCTGTGAAGAGACCTTC 1236
 QY 401 TyrThrMetLysValAsnAspGlyLysTyrValCysGluAlaAspLysPheTrpThrSer 420
 DB 1237 TACACAATGAAGTGAATGATGTAATATGTGTGAGGCTGATGATTCGTGAGACAC 1296
 QY 421 SerLysGlyLysSerLeuProValCysGluProValCysGlyLeuSerAlaArgThr 440
 DB 1297 TCCAAAGGAGAAAACATCCACAGCTGTGAGCCCTGTTGGACTATCAGCCCCACA 1356
 QY 441 ThrGlyLysArgIleTyrGlyGlnLysAlaLysProGlyAspPheProTrpGlnVal 460
 DB 1357 ACAGAGGCGCGTATATATGAGAGGCAAAAGCCAACTGGTATTTCTCTTGGCAAGTC 1416
 QY 461 LeuIleLeuGlyLysThrAlaAlaGlyAlaLeuLeuTyrAspAsnTrpValLeuThr 480
 DB 1417 CTGATATTGTAGTGGAACACACAGCAGCAGCTTTTATATGACAATGGGCTTACACA 1476
 QY 481 AlaAlaHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAspIleArgMetGly 500
 DB 477 GCTGCTCATGCCGTATGAGCAAAACATGATGATCGCCCTGGACATTCGAATGGGC 1536
 QY 501 ThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrpSerGluAlaValPheIleHis 520
 DB 1537 ACCCTGAAAAGACTACACCTCATTTACACAAGCTGGTCTGAACCTGTTTATATACAT 1596
 QY 521 GluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsn 540
 DB 1597 GAAGGTATTACTCATGATGCTGCTTGTGACAAATGACATGACTGATTAATGGAATAC 1656
 QY 541 LysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSer 560
 DB 1657 AAGATTGTAATCAATAGCAACATCAGCCTATTGTCTCCAAAGAAAGCTGAATCC 1716
 QY 561 PheMetArgThrAspAspIleGlyThrAlaSerGlyTrpGlyLeuThrGlnArgGlyPhe 580
 DB 1717 TTTATAGACACAGATGACTTGGAACTGCATCTGGATGGGATTAACCCAAAGGGCTTTT 1776
 QY 581 LeuAlaArgAsnLeuMetTyrValAspIleProIleValAspHisGlnLysCysThrAla 600
 DB 1777 CTGTGTAAGAAATCAATATGATGTGACATACCATGTTGACCATCAAAATGTACTGCT 1836
 QY 601 AlaTyrGluLysProProTyrProArgGlySerValThrAlaAsnMetLeuCysAlaGly 620
 DB 1837 GCATATGAAAGCCACCTATCCAAAGGGAAGTGAACGTCAACATGCTTGTCTGTGC 1896

QY 621 LeuGluSerGlyLysAspSerCysArgGlyAspSerGlyValAlaLeuValPheLeu 640
 DB 1897 TTGAAAGTGGGGGCAAGACACAGCTGCAGAGCGGAGGGGCACTGGTGTTCCTTA 1956
 QY 641 AspSerGluThrGluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsnLys 660
 DB 1957 GATAGTGAACAGAGAGGAGGTTTGTGGAGAGATAGTGTCTCGGGGTTCCATGAAATCT 2016
 QY 661 GlyGluAlaGlyLysTyrGlyValTyrThrLysValIleAsnTyrIleProTrpIleGlu 680
 DB 2017 GGGGAAGCAGCTCAGTATGAGTCTACACAAAGTATTAACTATATCTCGATCGAG 2076
 QY 681 AsnIleIleSerAspPhe 686
 DB 2077 AACATATTATGATTTT 2094
 RESULT 2
 HSMASP2PR 2455 bp mRNA linear PRI 12-DEC-1998
 LOCUS
 DEFINITION
 ACCESSION Y09926
 VERSION Y09926.1 GI:4007626
 KEYWORDS MASP-2 gene.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 AUTHORS Jensen, T.V.
 TITLE Direct Submission
 JOURNAL Submitted (06-DEC-1996) T.V. Jensen, University of Aarhus,
 Microbiology & Immunology, Bartholin Building, 8000 Aarhus C,
 DENMARK
 REMARK
 REVISION 2 (bases 1 to 2455)
 AUTHORS Thiel, S., Jensen, T.V., Stover, C.M., Schwaebler, W.J., Laursen, S.B.,
 Poulsen, K., Willis, A.C., Eggleton, P., Hansen, S., Holmskov, U.,
 Reid, K.B.M. and Jensenius, J.C.
 TITLE A second serine protease associated with mannin-binding lectin that
 activates complement
 JOURNAL Nature 386 (6624), 506-510 (1997)
 MEDLINE 97242412
 PUBMED 9087411
 REFERENCE
 AUTHORS Schwaebler, W.J.
 TITLE Direct Submission
 JOURNAL Submitted (08-DEC-1998) W.J. Schwaebler, Department of Immunology
 and Microbiology, University of Leicester, University Road,
 Leicester LE1 9HN, UK
 COMMENT
 FEATUSES
 source
 location/Qualifiers
 1..2455
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p36.2-3"
 /tissue_type="liver"
 1..2455
 /gene="MASP-2"
 1..16
 /gene="MASP-2"
 17..2077
 /gene="MASP-2"
 /codon_start=1
 /evidence="experimental"
 /product="MASP-2 protein"
 /protein_id="CAA71059.1"
 /db_xref="GI:4007627"
 /db_xref="SPTREMBL:O00187"
 /translation="WRLLTLGLGLGSAVATPLGPKWPEPVFGRLASPGFPGPEYANDQE
 RRMVLTAPPGRRLRLFTFHFDELSLCEYFVKLSGAKVLAATLGGESDTERAPG
 KDTFYSLSDDITFRSDYSNEKPTGFEAFYAABDIDECQVAPGEAPTCDDHCHNHL
 GGFYCSCRAGYLVHRNKRCSALSGGVFTQRSGLSSPEVPRPRPKLSSQYISLSE


```

|||||
Db 1757 CTGTGAGAAATCTATGATGCGACATACCGATTGTTGACCATCAAAATGACTGCT 1816
QY 601 AATyGluLySerProArgGlySerValThralasmetLeuCySalaGly 620
Db 1817 GCATATGAAAGCCACCCTATCCAGGGGAGGTACTGCTACATGCTTGTGCTGGC 1876
QY 621 LeuGluSerGlyLySerSerCySArgGlyAspSerGlyLyAlaLeuValPheLeu 640
Db 1877 TTAGAAAGTGGGGGCAAGGACCTGACAGGTGACAGCGGAGGGGCGACTGCTTCTTA 1936
QY 641 AspSerGluThrGluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsnCyS 660
Db 1937 GATAGTGAACACAGAGAGTGTGTTGGGAGGAATAGTGTCTGGGGTTCATGAATTGT 1996
QY 661 GlyGluAlaGlyLyGlyTrpGlyValTrpThrLyValIleAsnTrpIleProTrpIleGlu 680
Db 1997 GGGGAGCGAGGTAGTATGAGGTCTACACAAAGTTATTAATATTCCTGATCGAG 2056
Db 2057 AACATAATTAGTATTTT 2074

RESULT 3
HOSAI8286 4464 bp DNA linear PRI 12-JUL-1999
LOCUS Homo sapiens partial MASP-2 gene, clone pgm-2A.
DEFINITION Y18286
ACCESSION Y18286
VERSION 1
KEYWORDS GI:5459322
SOURCE lectin-associated protein; mannose binding lectin-associated
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 4464)
Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,
Jensenius,J.C. and Schwaebble,W.J.
Two constituents of the initiation complex of the mannan-binding
lectin activation pathway of complement are encoded by a single
structural gene
J. Immunol. 162 (6), 3481-3490 (1999)
JOURNAL 99192764
MEDLINE 10092804
PUBMED 2 (bases 1 to 4464)
REFERENCE Schwaebble,W.J.
AUTHORS Direct Submission
TITLE Submitted (19-Oct-1998) W.J. Schwaebble, Department of Microbiology
and, Immunology, University of Leicester, University Road, PO Box
138, Leicester LE1 9HN, UK
FEATURES
source
1..4464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2-3"
/clone="pgm-2A"
/cell_type="lymphocyte"
1..4464
/gene="MASP-2"
join(1..559,2414..2610,2926..4464)
mRNA
/gene="MASP-2"
/gene="alternative"
join(1..559,1004..1156)
mRNA
/gene="MASP-2"
/gene="alternative"
1..559
exon
/gene="MASP-2"
/note="a"
join(16..559,2414..2610,2926..4245)
CDS
/gene="MASP-2"
/note="alternative"
/codon_start=1

```

```

/product="MASP-2 protein"
/protein_id="CAB50733.1"
/db_xref="GI:5459324"
/translation="MKLLTLGLLGGSVATPIGPKMPPEVFGRLASPGFPGRYANDOE
RMWTLAPPGYRLRYLTFHFDLLESHLCYDVKLSGAKVLATLGGESTDTERAPG
KDTFVSLGSSLDITFRSDYSNEKPTGEFAFAADIDECQVAPEDAPPCDHNHNL
GGFYCSRAGYVLHNRNKRITCSALCSGQVOTQSGEISPEYRPYPKLSSTCYSTLE
EGFSYILDFVESFDEVTHEPTLCSPYDFELIOTDREHSGPFGKTLPHRIETSNVTI
TFPTDESGDHTGCMKIHVTSSTAHACYPMAAPMGHVSPOAKYLKDSPIFCEYVEL
LOGHLPLKSFATVCKDGSMDRPMALYDCCGPPDOLCSGSEVEYITTGCGVTTAAVI
QTSCEFTYTMKVNDEKTYVEADGFWTSKESLSLPEVCGLSARFTGGKTYGGQK
AKPGDFPMOVLGGTTAAAGALLILDNMWVLTAAHAYVEQNDASALDIRGTLKRLSPH
YTOAMSEAVFIHEGYTHDGFNDIALILNKNVYINSNITPICPRKAESEFRTDD
IGTASGMLGTORGFLARNLMYVDIPIVDHOKCTAAVEKPYRGSVTAMMLCAGLESG
GKDSRGDGGGALVFLDSEFTERFWFGIYSMSNMNGEAGOVYVTKVINYIIPNIENI
ISDP"
join(16..559,1004..1017)
/gene="MASP-2"
/note="alternative"
/codon_start=1
/product="MAP19 protein"
/protein_id="CAB50732.1"
/db_xref="GI:5459323"
/translation="MKLLTLGLLGGSVATPIGPKMPPEVFGRLASPGFPGRYANDOE
RMWTLAPPGYRLRYLTFHFDLLESHLCYDVKLSGAKVLATLGGESTDTERAPG
KDTFVSLGSSLDITFRSDYSNEKPTGEFAFAADIDECQVAPEDAPPCDHNHNL
GGFYCSRAGYVLHNRNKRITCSALCSGQVOTQSGEISPEYRPYPKLSSTCYSTLE
EGFSYILDFVESFDEVTHEPTLCSPYDFELIOTDREHSGPFGKTLPHRIETSNVTI
TFPTDESGDHTGCMKIHVTSSTAHACYPMAAPMGHVSPOAKYLKDSPIFCEYVEL
LOGHLPLKSFATVCKDGSMDRPMALYDCCGPPDOLCSGSEVEYITTGCGVTTAAVI
QTSCEFTYTMKVNDEKTYVEADGFWTSKESLSLPEVCGLSARFTGGKTYGGQK
AKPGDFPMOVLGGTTAAAGALLILDNMWVLTAAHAYVEQNDASALDIRGTLKRLSPH
YTOAMSEAVFIHEGYTHDGFNDIALILNKNVYINSNITPICPRKAESEFRTDD
IGTASGMLGTORGFLARNLMYVDIPIVDHOKCTAAVEKPYRGSVTAMMLCAGLESG
GKDSRGDGGGALVFLDSEFTERFWFGIYSMSNMNGEAGOVYVTKVINYIIPNIENI
ISDP"
CDS
join(16..559,1004..1017)
/gene="MASP-2"
/note="alternative"
/codon_start=1
/product="MAP19 protein"
/protein_id="CAB50732.1"
/db_xref="GI:5459323"
/translation="MKLLTLGLLGGSVATPIGPKMPPEVFGRLASPGFPGRYANDOE
RMWTLAPPGYRLRYLTFHFDLLESHLCYDVKLSGAKVLATLGGESTDTERAPG
KDTFVSLGSSLDITFRSDYSNEKPTGEFAFAADIDECQVAPEDAPPCDHNHNL
GGFYCSRAGYVLHNRNKRITCSALCSGQVOTQSGEISPEYRPYPKLSSTCYSTLE
EGFSYILDFVESFDEVTHEPTLCSPYDFELIOTDREHSGPFGKTLPHRIETSNVTI
TFPTDESGDHTGCMKIHVTSSTAHACYPMAAPMGHVSPOAKYLKDSPIFCEYVEL
LOGHLPLKSFATVCKDGSMDRPMALYDCCGPPDOLCSGSEVEYITTGCGVTTAAVI
QTSCEFTYTMKVNDEKTYVEADGFWTSKESLSLPEVCGLSARFTGGKTYGGQK
AKPGDFPMOVLGGTTAAAGALLILDNMWVLTAAHAYVEQNDASALDIRGTLKRLSPH
YTOAMSEAVFIHEGYTHDGFNDIALILNKNVYINSNITPICPRKAESEFRTDD
IGTASGMLGTORGFLARNLMYVDIPIVDHOKCTAAVEKPYRGSVTAMMLCAGLESG
GKDSRGDGGGALVFLDSEFTERFWFGIYSMSNMNGEAGOVYVTKVINYIIPNIENI
ISDP"
BASE COUNT 938 a 1338 c 1217 g 971 t
ORIGIN
Alignment Scores:
Pred. No.: 4,56e-253 Length: 4464
Score: 3298.50 Matches: 680
Percent Similarity: 48.12% Conservative: 0
Best Local Similarity: 48.12% Mismatches: 2
Query Match: 87.77% Indels: 731
DB: 9 Gaps: 6
US-09-874-198-2 (1-686) x HOSAI8286 (1-4464)
QY 1 MetArgLeuThrLeuGlyLeuLeuCySGLySerValAlaThrProLeuGlyPro 20
Db 16 ATGAGGCTGCTGACCTCTCGGGCGCTTCTGTGGCTGGTGCCACCCCTTAGGCGCG 75
QY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAla 40
Db 76 AAGTGGCTGAACCTGTGTGGGGCGCTGCATCCCGCGCTTTCAGGGAGTATGCC 135
QY 41 AsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
Db 136 AATGACCAGAGGAGCGGCTGACCTGACCTGACACCCCGCGCTTTCAGGCGCTTAC 195
QY 61 PheThrHisPheAspLeuGluLeuSerHisLeuCySGLyTrpAspPheValLysLeuSer 80
Db 196 TTACACCACTTGACCTGGAGCTCTCCACCTTGCAGAGTACACTTGCAGAGCTGAGC 255
QY 81 SerGlyAlaLysValLeuAlaThrLeuCySGLyGlnGluSerHisAspTrpGluArgAla 100
Db 256 TCGGGGGCCAAAGTGTCTGCGCAGCTGTGGGGCAGAGACACAGACGAGCGGCGC 315
QY 101 ProGlyLysAspThrPheTrpSerLeuGlySerSerLeuAspIleThrPheArgSerASP 120
Db 316 CTGTCGAAGACACTTCTACTGCTGGGCTCCAGCTGCATTTACCTTCGCTCGAC 375

```

Qy	121	TyterlsnqilulyrsPvropethnqilypheglualaIphetYrAlaIalaglsupllesp	140
Db	376	TACTCAACGAGAAGCGGTTCCAGGGGgTTCGAGGCTCTCATGCAAGCCAGACATTGAC	435
Qy	141	glucysgluValaIalaproglIyglualaIaproThCyAspHisHiscysHisnHisIeu	160
Db	436	GAGTGCACAGGTGGGCCCGGGAGAGGGCCCACTGGGACACCACTGCCACACCACTG	495
Qy	161	glyelYpethYcysSerCySargalaglyTyTrValIeuHisArgnulsYsrgrThCys	180
Db	496	GGCGGTTTCTACTGCTCTCTCCGCGCAGAGCTACGTCTCGACCGTACACAGGCACCTGC	555
Qy	181	Ser-----	181
Db	556	TCAGGTGAGGAGAGCTGCCTGGGCCCCACGACCCCTCTCTGGGATPACCGGGCTCCT	615
Qy	181	-----	181
Db	616	CAGGGCCATTGCTGCTCTGTCCCAAGGGTGCGGAGGGCTTGGGCTTGGACATGGGTGTT	675
Qy	181	-----	181
Db	676	CTAAGCCCTGTGCTGCCTGCACAGTCCCTCTTCACGCCCTGGTTCCCTCTCAGCAGCCAGGC	735
Qy	181	-----	181
Db	736	TCATCAGTGCACCCCTGCCTGACACTGAGACTAATTTAAGATCCCACTGTACTGG	795
Qy	181	-----	181
Db	796	TTCCACCTGGGCTCTGGGAACCCCTCATGTAGCCACGGAGAGTGGGGTATCTAGCTC	855
Qy	181	-----	181
Db	856	GTTCTTGGACTGTTCCTGTTCCTGTGACTGTGGACAGGCCAGTGTCTGGAGCTGGG	915
Qy	181	-----	181
Db	916	CAGCCCAACCTGTGGGCTGACCTGCCTCCCGACTCGGTTTCTCCTCGGTGTCT	975
Qy	181	-----	181
Db	976	CTTCCTTGCCTCTGTGATCTCTTTTCCAGACAGAGCCTTACGCTCCCTGSAGCTCCGG	1035
Qy	181	-----	181
Db	1036	CCTGCCACAGCTAGAGAAAGCCAGAGCCAGCTGCTGTACGCTCCGGSTTGGGTGAGA	1095
Qy	181	-----	181
Db	1096	TGGCTGTGCCCACTCCCATTCACCAACATGAGACCAATTAATAACGTGCCCCACCCC	1155
Qy	181	-----	181
Db	1156	ACCTGTGCAGCTGTCTCTGTGGGgTGGAGGGTCTGGGAGCGGGGTGCGTCTCTCTC	1215
Qy	181	-----	181
Db	1216	TGCTACCCCTCTCACAGCCTCATGAACCCACAGCTGTGTGGAGCTCTCATGGGGCC	1275
Qy	181	-----	181
Db	1276	ACACGGTCTTGGGCTACACCCCTGTTTTTGAAGATGGGCACTGAGGCGCGAGAGGGTA	1335
Qy	181	-----	181
Db	1336	AGGCGTCGCTGAGTCCAGAGTCCCCAGAGGCTGAGCCAGAGTAACTTGAACCAACCCC	1395
Qy	181	-----	181
Db	1396	ATTCAAGGCTTGGGCTTGGAGAGCCTGACCCACAGAGGAGACACCTTGGAGATTTTCAT	1455
Qy	181	-----	181

Db	1436	TGAGGGGTAATCTGTGTCCCGCAAAATCCAGGGGTGATTGCCACTGCCATCGCCCATAGGCACAG	1513
Qy	181	-----	181
Db	1516	CCAGCTGGAGAGAGCAGGCAATGTTGGGGCTCTCCTACTTCTTAGAGGCTCACAACATCA	1575
Qy	181	-----	181
Db	1576	AATGCCCCCACTGCAGCTGGGGGTGGGTGTGTGATGGGATGGGGACCAAGCCTTCT	1635
Qy	181	-----	181
Db	1636	TGAAGATTAGAGCCCAAGCACCACCCCGCCGTGGCAGACAGCATCAGCTTTCACAG	1699
Qy	181	-----	181
Db	1696	AGGAAGAGAGCACCAGACACACATCATGATGATCATTGTGCTTGAACCTTCAGAGACAGCC	1759
Qy	181	-----	181
Db	1736	CCAGGGCAGGGGTACAGACTGGGGAAAGGGGTGATAGAGACCTTCTTCCGATGTTTC	1815
Qy	181	-----	181
Db	1816	CTCCAGAGACCAGGGGGCTGGCTGTCTTGGCTTGGGTGGGTAGAGAGACCCCTGATGAA	1875
Qy	181	-----	181
Db	1876	TAAACTTGGGAATCATCTGGGGTGGCTGTAGAGGAATTTAGGGAGCTCCGAAGGGGCCCT	1935
Qy	181	-----	181
Db	1936	TAGGCTGAGAGATGCTCTCTCTTTTCCCGAATTTCCAGGGACCAGAGAGATGTCCC	1995
Qy	181	-----	181
Db	1996	TTCTTCCCTCTCTGTGTGTCCATCCACCCCGCCCGCCCGCTGGCAGAGCTGTGTGA	2055
Qy	181	-----	181
Db	2056	ACTCAGTCTCTAGCCCTTACCCTGGGGTTTGGACTGTGGCTCAGAGATACACACAGCTTC	2115
Qy	181	-----	181
Db	2116	CTGGGGGTGATGAGAGGGGCTGTGGCTCCATCCCGAGTCTGTCTTTCACGTAAA	2175
Qy	181	-----	181
Db	2176	GCTTCAAAGCAAGAAACCCCTCTCTAAGGGGCCCTCAGACATCGGGTGGTCTTTT	2235
Qy	181	-----	181
Db	2236	GTTTCTGGGTAGGCTTCAGGGCTGGCCACTGCAGGGCCAGCCACAGGAGATCA	2295
Qy	181	-----	181
Db	2296	GATGTCCAGCCACATCCCTGTCCCACTTCTGTCTCCCAAGCATCCACCTCTCTTT	2355
Qy	182	-----Ala 182	
Db	2356	GGTGGAGAGGCTGTAGAGGGCAAGCCAACTACCTCCCGCCCTTCCCTCTCCAGGC	2415
Qy	183	LeucysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGluTyrPro	202
Db	2416	CTGTGCTCCGGCCAGGCTTCTCACCCAGAGGTCTGGGGAGCTCAGCAGCCCTGAATAACCA	2475
Qy	203	ArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGluIlePheSer	222
Db	2476	CGGCGGATCCCAAACTCTCAGTTGCACTTACAGCATAGCCTGGAGAGGGGTTCACT	2533
Qy	223	ValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeuCysPro	242

Db 2536 GTCATTCGTGACCTTTGTGAGTCTTCGATGTGGAGACACCCCTGAACCCCTGTGCC 2595
 Oy 243 TTAAspPheLeuLys----- 247
 Db 2596 TAGGACTTTCTCAAGGCTGTGGCTCTGGCCCCCTCATCTTGTCCAGATCCTCCCCCTTC 2655
 Oy 247 ----- 247
 Db 2656 AGCCAGCTGCACCCCTACTTCTGTGAGCATGGCCCCCACAACGTTCCGGTCAACCTCG 2715
 Oy 247 ----- 247
 Db 2716 GTGACCCCACTCTTCAGTGTCTATGAGAGTCAAGGCTGGGGCTTCGAGTACAAAGTG 2775
 Oy 247 ----- 247
 Db 2776 TGGGAGGACAGTGGGGAGGGGACCCCAATCCATGGCTGGGTGGCTCATTTGGCTGT 2835
 247 ----- 247
 836 CCCTGAATGCTGAGAGGTGGGTACTTCCCTCCGCCAGGCCAGACCCAGCAGCTGC 2895
 Oy 248 ----- 248
 Db 2896 TTCCCACTTTCATGAGCTTCTTCTCATAGATTCAAACAGACAGAGAAACATGGCCCA 2955
 Oy 258 PhcCysGlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThr 277
 2956 TTCTGTGGAGACATTTGCCCAAGATGAAACAAAGACACAGGAGCATTACCATACC 3015
 Oy 278 PheValThrAspGluSerGlyAspHisThrGlyTrpLysIleHisThrSerThrAla 297
 3016 TTGTGCACAGATGATCAGAGACAGACACAGGCTGAAGATCCACACAGAGACAGCGC 3075
 Db 298 GluProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLys 317
 3076 CACGCTTGCCCTTATCGAATGGGGCCCACTTAATGGCCACTTTCACCTGTGCAAGCCAAA 3135
 Db 318 TyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTrpGluLeuLysGly 337
 3136 TACATCTCTGAAGACAGACTTCTCATCTTTTGGCAGACTGGCTATGAGCTTCTGCAGGT 3195
 Oy 338 HisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspArgPro 357
 3196 CACTTGCCCTGAATCTCTTACTGACGTTGTGCAGAAAGATGATCTTGGAGCCGCCA 3255
 Db 358 MetProAlaLysSerIleValAspCysGlyProProAspAspLeu-----ProSerGly 375
 256 ATGCCCGCG-----ATTGTTGACTGTGGCCCTCTCATGATCATCCAGCCCGAGTGGC 3309
 Oy 376 ArgValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSer 395
 3310 CGAGTGGAGTACATCACAGGTCTCTGGAGTACACACTACAAAGCTGTGATTCACATACGC 3369
 Oy 396 CysGluGluThrPheTyrThrMetLysValAsnAspGlyLysTyrValCysGluAlaAsp 415
 3370 TGGAGACAGACTTCTACACAAATGAAGTGAATGATGTAATATGTGTGAGCGCAT 3429
 Db 416 GlyPheThrPheSerLysGlyGlyLysSerLeuProValCysGluProValCysGly 435
 3430 GGAATTCGGACGAGCTCCAAAGGAAATACTCCAGCTGTGTGAGCCGTGTTGTGGA 3489
 Oy 436 LeuSerAlaArgThrThrGlyGlyArgIleTyrGlyGlyLysAlaLysProGlyAsp 455
 3490 CTATCAGCCCGCACACAGAGGCGGTATATATGAGAGCAAAAGGCAAAACCTGGTAT 3549
 Oy 456 PheProTrpGlnValLeuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeu-----Leu 473
 3550 TTTCCTTGGCAAGTC-----TTAGTGGAAACACACAGAGGTGCACATTCGATTTA 3603
 Oy 474 TyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysHisAspAlaSer 493
 3604 TATGACAACGTGGTCTTAACAGCTGTATGAGCCGTATATGAGCAAAACATGATGATCC 3663

Oy 494 AlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrp 513
 Db 3664 GCCCTGGACATTTGGAATGGCACCCTGAAGACTATACCTCTTTATACAAAGCTTGG 3723
 Oy 514 SerGluAlaValPheIleHisGlyTyrThrHisAspAlaGlyPheAspAspIle 533
 Db 3724 TGTGAAGCTGTTTTATACATGAAGGTATATCATCATGATGCTGGCTTGGCAATGACATA 3783
 Oy 534 AlaLeuIleLysLeuAsnLysValValIleAsnSerAsnIleThrProIleCysLeu 553
 3784 GCACGTAAATTAATTGAATTAACAAAGTTGTATCATATGCAACAATCACGCTATTGTCTG 3843
 Oy 554 ProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSerGlyTrp 573
 Db 3844 CCAGAAAGAGAGCTGATCTTATATGAGACAGATCATGTGAGCTGATCGATGG 3903
 Oy 574 GlyLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMetTyrValAspIleProIleVal 593
 Db 3904 GGATTAACCCAAAGGGGTTTTCTTGTGTAAGAAATCTATATGATGACATACCATTTGTT 3963
 Oy 594 AspHisGlnLysCysThrAlaAlaTyrGlyLysProProTyrProArgGlySerValThr 613
 Db 3964 GACCATCAAAATGCTACTGCTGCATATGAAGCCACCTATCCAAAGGGAGGTACT 4023
 Oy 614 AlaAsnMetLeuCysAlaGlyLeuGluSerGlyLysLysAspSerCysArgGlyAspSer 633
 Db 4024 GCTAACCTGCTTTGTCTGCTTGAAGATGGGGCAAGACACTGCAGAGGTGACAGC 4083
 Oy 634 GlyGlyAlaLeuValPheLeuAspSerGluThrGluArgTrpPheValGlyIleVal 653
 Db 4084 GGAGCGGACACTGGGTTCATAGATAGTGAACAGAGAGGTGGTTGGAGGAATAGTG 4143
 Oy 654 SerTrpGlySerMetAsnCysGlyGluAlaGlyLysIleGlyValTyrThrLysValIle 673
 Db 4144 TCCGTGGGTTCATGAATTTGGGGAAGCAGGTCTAGTATGAGCTACACAAAGATTATT 4203
 Oy 674 AsnTrpIleProTrpIleGluAsnIleIleSerAspPhe 686
 Db 4204 AACTATTTCCCTGATCGAAGACATTAATTAATTTT 4242

RESULT 4
 HOSAI8287 4465 bp DNA linear PRI 12-JUL-1999
 LOCUS Homo sapiens partial MASP-2 gene, clone pgm-2B.
 DEFINITION Y18287.1 GI:5459325
 ACCESSION Y18287.1
 VERSION
 KEYWORDS lectin-associated protein; mannose binding protein; serine protease.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 4465)
 AUTHORS Soveri,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T., Jensenius,J.C. and Schwaebble,W.J.
 TITLE Two constituents of the initiation complex of the mannan-binding lectin activation pathway of complement are encoded by a single structural gene
 JOURNAL J. Immunol. 162 (6), 3481-3490 (1999).
 MEDLINE 99192764
 PUBMED 10092804
 REFERENCE 2 (bases 1 to 4465)
 AUTHORS Schwaebble,W.J.
 TITLE Direct Submission
 JOURNAL Submitted (19-OCT-1998) W.J. Schwaebble, Department of Microbiology and, Immunology, University of Leicester, University Road, PO Box 138, Leicester LE1 9HN, UK
 FEATURES
 source
 1..4465
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"

QY 180 ----- 180
Db 1276 CACACGGTCTTGCCCTCAACCCCTGTTTGAAGATGGGGGACTGAGGCCGAGAGGGGT 1335
QY 180 ----- 180
Db 1336 AAGCCTCGCTCGAGTCCAGGTCCAGAGAGGTAGCCAGAGTAATCTTGAACACCCC 1395
QY 180 ----- 180
Db 1396 CATTCAGAGGTCTGGCTTGAGAGCCTGACCCACAGAGAGACCCCTGGAGATATTC 1455
QY 180 ----- 180
Db 1456 TTGAGGGGTAACTGTCGCCCGCAATCCAGGGGTGATTCGCCAGTCGCCCATAGCACA 1515
QY 180 ----- 180
Db 1516 GCCACGTGGAAGAAGCAGCAGCAATGTTGGGGTCTCTCACTTCTAGAGGCTCACAAC 1575
QY 180 ----- 180
Db 1576 AAATGCCCCCTGAGCTGAGGTTGGGGTGGGTGGTATGGGATGGGACCAAGCCTTCC 1635
QY 180 ----- 180
Db 1636 TTGAAGATAGAGCCAGCCAGCCAGACCCCGCCCGTGGCAGACATCAGGTGTTCCAGC 1695
QY 180 ----- 180
Db 1696 GAGGAAGAGAGCAGCAGACTCAGTCATGATCACTGTTGCCCTGAACCTCCAGAGCAGC 1755
QY 180 ----- 180
Db 1756 CCCAGGCGAGGGGTCAAGACTGAGCTGAGGAAAGGGGTGATGAGAGACCCTTCTCCGAGATT 1815
QY 180 ----- 180
Db 1816 CCTCAGGAACACAGGGGCTGGGCTGTGGCTTGGGTTGGGTAGGAGACCCCTGATGA 1875
QY 180 ----- 180
Db 1876 ATAACTTGGGAATCACTGGGGTGGCTGTAAAGGAATTTAGGGAGCTCCGAAGGGCCC 1935
QY 180 ----- 180
Db 1936 TTAGGCTCGAGAGATGCTCTCTCTTTTCCGAATTTCCAGGAGCAAGAGAGTGTCC 1995
QY 180 ----- 180
Db 1996 CTTCCTTCCTCTCTGTGTGTCATCCACCCCGCCCGCCCGCTGGCAGAGCTGCTGG 2055
QY 180 ----- 180
Db 2056 AACTCAGTCTTAGCCCTACCCCTGGGGTGGAGACTGTGCTCAGAGATCACACAGCT 2115
QY 180 ----- 180
Db 2116 CTTGGGGGTGTGAGTAGAGGGGCTGTGCGCTCATCCGAGTGTGCTGCTTTTCAAGTAA 2175
QY 180 ----- 180
Db 2176 AGCCTCAAGCAGAGAAACCCCTCTCTAAGGGGCCCCCTCAGCCATCGGGTGGGTGTT 2235
QY 180 ----- 180
Db 2236 TGGTTTCTGGGTAGGCTCAGAGGGGTGGCCACCTGCAGGGGCCAGCAACCAAGGATGC 2295
QY 180 ----- 180
Db 2296 AGATGTCACGACATCCCTGTGCCAGTTTCTGCTCCCAAGGATCCACCCCTGCTGT 2355

QY 181 -----Ser-A1 182
Db 2356 TGGTGGAGGGGCTGATAGAGGAGCGCAAGTCACTCCCTGGCCCTTCCCTCTTCACG 2415
QY 182 aLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerProGluTyrPr 202
Db 2416 CTTGTGCTCCGGCAGGCTTTCACCCAGAGAGTCTGGGGAGCTCAGAGCCCTGAATACCC 2475
QY 202 oArgProTyrProLysLeuSerCysThrTyrPheSerLeuGluGluLysPhe 222
Db 2476 ACGGCGTATCCCAACTCTCCAGTTGCAAGCATAGCATAGCCTGGAGGAGGCTTCAG 2535
QY 222 rValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeuCysPr 242
Db 2536 TGTCAATCTGTGACTTGTGTGAGTCTTCGATGTGGAGACACACCCTGAACCTGTGTCC 2595
QY 242 oTyrAspPheLeuLys----- 247
Db 2596 CTACGACTTTCACAGGTCTGCTGCTGGGCCCTCATCTTGTCCAGATCTCCCTT 2655
QY 247 ----- 247
Db 2656 CAGCCAGCTGCACCCCTACTTCTGACAGATGCCCCCAGCAGCTTCCGTCACCTTC 2715
QY 247 ----- 247
Db 2716 GTGACCCCACTCTTCAGGCTCTATGAGGTCAAGGCTGGGCTTGAGTACACT 2775
QY 247 ----- 247
Db 2776 GTGGAAGCAGAGTGGGAGGGGACCCCAATCATGCGCTGGGTGGCTCATTTGGCTG 2835
QY 247 ----- 247
Db 2836 TCCCTGAATGCTGAGAGGTGGTTACTTCCCTCCGCCAGCCAGACCCAGCAGCCTG 2895
QY 248 -----IleGlnThrAspArgGluGlnHisGlyPr 257
Db 2896 CTTCGCCAGCTTCATGAGCTTCTTCTCAGATTTCAAAACAGACAGAGAACATGGCCC 2955
QY 257 oPheCysGlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleTh 277
Db 2956 ATTCTGGGAAACATTTGCCCAAGAGATTGAACAAAGAACACCGGTGACCATCAC 3015
QY 277 rPheValThrAspGluSerLysPheHisThrGlyTyrPheLysThrSerThrAl 297
Db 3016 CTTTGTACAGATGAATCAGAGACCAACAGGCTGGAAATGCCACTACACGAGCAGC 3075
QY 297 aGlnProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLys 317
Db 3076 GCACGCTTGGCTTATCCGATGGCGCCACCTAATGGCCAGCTTTCACCTGTGCAGGCC 3135
QY 317 sTyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnI 337
Db 3136 ATACATCTGAAGACAGCTTCTTCATCTTTCGAGACAGCTGATAGCTTTCGCAAGC 3195
QY 337 yHisLeuProLeuLysSerPheThrAlaValLysGlnLysAspGlySerThrPaspArgPr 357
Db 3196 TCACCTTGGCCCTGAATACCTTCTGACGTTGTGCAGAAAGATGAGATCTTGGAGCCGCC 3255
QY 357 oMetProAlaCysSerIleValAspCysGlyProProAspAspLeu-----ProSerG1 375
Db 3256 AATGCCCGCG-----ATTGTGACTGTGGCCCTCTCGATGATCTATGACGCCCACTGG 3309
QY 375 yArgValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSe 395
Db 3310 CCGAGTGAAGTACATCAGAGTCTCTGGAGTGCACACTACAAAGCTGTGATTCAGTACAG 3369
QY 395 rCysGluGluThrPheThrThrMetLysValAsnAspGlyLysTyrValCysGlnAlaAs 415
Db 3370 CTGTGAAGAGACCTTCTTACACATGAAGATGAAGTGAATATGTGTGTGAGCTGTGA 3429
QY 415 pGlyPheTrpThrSerSerLysGlyGluLysSerLeuProValCysGluProValCysG1 435

```

Db 3430 TGGATTCTGGACGAGCTCCAAAGAGAAAATCACTCCAGCTGCTGTGAGCTGTTGTGG 3489
QY 435 yLeuSerAlaArThrThrlGlyArgIleTyrglyGlyGlnLysAlaLysProGlyAs 455
Db 3490 ACATATGAGCCCGACAAACAGAGGGCGCTATATATGAGGGCAAAAGCAAACTGGTGA 3549
QY 455 pPheProTrGlnValLeuIleLeuGlyGlyThrThAlaAlaGlyAlaLeu-----Le 473
Db 3550 TTTTCTGGCAAGTC-----TTAGTGGACACACAGCAGGTGCATTCTCGAATTT 3603
QY 473 uTyraSPasTrPValLeuThrAlaAlaHisAlaValTyrglyGlnLysHisaspAlaSe 493
Db 3604 ATATGACAAATGGGCTCTTAACACCTGCTCATGCCGTCTATGAGCAAAACATGATGCATC 3663
QY 493 fAlaLeuAspIleArgMetIlyThrLeuLysArgLeuSerProHisTyrrhrGlnAlaTr 513
Db 3664 CGCCCTGGACATTGGAATGGGACCCCTGAAAGACTATACCTCATTTATACACAGCCTG 3723
QY 513 pSerGluAlaValPheIleHisGluGlyTyrrhrHisaspAlaGlyPheaspAsnAspII 533
Db 724 GTCTGAAGCTGTTTATATACATGAGGTATACTCATGATGCTGCTTGACAAATGCAT 3783
QY 533 eAlaLeuIleLysLeuAsnAsnLysValValIleasnSerAsnIleThrProIleCysLe 553
Db 3784 AGCAGCTGATTAATTCATTAACAAGCTTGAATCAATACACATCACGCCCTATTGTCT 3843
QY 553 uProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSerGlyTr 573
Db 3844 GCCAGAGAAAAGAGCTCAATCTTTATGAGACAGATGACATTTGAAGCTCATCTGGATG 3903
QY 573 pGlyLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMetTyrrValAspIleProIleVa 593
Db 3904 GGATTAACCCAAAGGGGTTTCTGCTGCTAAGAAATCTAATGATGTCACATACCGATGT 3963
QY 593 lAspHisGlnLysCysThrAlaAlaTyrglyLysProProTyrrProArgLysSerValTh 613
Db 3964 TGACACATCAAAAATGTACTCTCATGTGAAAAGCCACCTATCAAGGGAGAGTGAAC 4023
QY 613 fAlaAsnMetLeuCysAlaIleGluSerGlyGlyLysAspSerCysArgGlyAspSe 633
Db 4024 TGCATACATCTTGTGCTGCTGCTAGAAAGTGGGGGACAGACCTGCAGAGGTGACAG 4083
QY 633 rGlyGlyAlaLeuValPheLeuAspSerGluThrGlnArgTrPheValGlyIleVa 653
Db 4084 CGAGGGGGCCTGCTGTTCTTACATAGTGAACAGAGAGGTGTTCTGGAGCAATAGT 4143
QY 653 lSerTrpGlySerMetAsnCysGlyGluAlaGlyGlnTyrglyValTyrrhrLysValII 673
Db 1144 GTCCCTGGGGTTCATGATGATGTGGGAGAGGTCAGTATGAGTACACAAAAGTTAT 4203
QY 673 eAsnTyrrIleProTrIleGluLysnIleIleSerAspPhe 686
Db 4204 TAACATATTCCTCGATGAGAACATTAATTAGTGAATTTT 4243

```

RESULT 5
BC013893
LOCUS BC013893 3070 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, Similar to mannan-binding lectin serine protease 2,
clone MGC:13718 IMAGE:4211293, mRNA, complete cds.
ACCESSION BC013893
VERSION BC013893.1 GI:15530224
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadams@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRK Plate: 18 Row: n Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6754641.
Location/Qualifiers
1. 3070
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:13718 IMAGE:4211293"
/tissue_type="Liver, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B"
/note="Vector: pCW-SPORT6"
17. .2074
/codon_start=1
/product="Similar to mannan-binding lectin serine protease
2"
/protein_id="AAH13893.1"
/db_xref="GI:15530225"
/translation="MRLLFLGLMSVATLLGSKWPPVFRGLYSPPREKYADHD
RSMVTLAPPGYRLRYFTHEDELSEYRCEDYDVAKLSGKVLATLCGDESTDGAPG
NDTFYSLGSLVTFHSDYSNEKPTGFEEAFVADVDCEKRVSLDSDYCHYCHNL
GGYCSRAGVYLHQNKHTCSALCGOVTGSGYLSSEYPOQPKLSSCYSTRLE
DGEFVILDFVESFEDVETHPACCPDLSLKTORDKEHGPFGCTLPRIETMSHYTI
TRATDESKNHTKMKIHYTSTARPCDPPAPRPGCSISPOAIVYILDRSEVCKTPEL
LOGSVPLKFTAVQCDQSDWDRMPBCSLTIDCGPPDDLPHGVDTITTEVNTTYAVI
QYSCETEFYVNSNKKYCEADGFWTSSGKGLPVCEVCGLSHTTGGRTVGQPA
KPGDEPMQVLLLGOTTAAGALIDHNWVLTAAHAYEKRMASLIRMGILKRSPL
YQAAPEELFIEGTYHGAFNDIALILKRVKIVNGSIMPVCLPRKEASLMTDF
TGTAVGMDGLTGKGLARNLMPVDIPIADHOKCTAYEKIYPCGVRSANMLCGLEGG
KDSKRDSCGALVFLDNETGRKRFVGGIYVMSGINSINGADQIVYTVKIVYIWMENII
SNP"

BASE COUNT 836 a 756 c 708 g 770 t
ORIGIN

Alignment Scores:
Pred. No.: 1 06e-234 Length: 3070
Score: 3065.50 Matches: 560
Percent Similarity: 88.65% Conservative: 49
Best Local Similarity: 81.51% Mismatches: 75
Query Match: 81.57% Indels: 3
DB: 10 Gaps: 3

US-09-874-198-2 (1-686) x BC013893 (1-3070)

QY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuCysGlySerValAlaThrProLeuGlyPro 20
|||||
Db 17 ATGAGGCTACTCATCTTCCTGGGCTCGTGAGTGTGGTGGCCACACTTCGGGTGCA 76
|||||
QY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyTyrrAla 40
|||||
Db 77 AAGTGGCTGAACCTGTATTCCGGGCCCTGTGTTCCCTGCTGCCACAGAGATGTGCT 136
|||||
QY 41 AsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyLysTyrrArgLeuArgLeuTyrr 60
|||||
Db 137 GACCATCAGATGATCCTGGACATGACATGCCCCCTGGCTACCCGCTGCCCTCTAC 196
|||||
QY 61 PheThrHisPheaspLeuLeuLeuSerHisLeuCysGlyuTyrrAspPheValLysLeuSer 80
|||||

```

Db 197 TTACCCACCTTTGACCGGAACCTCTTACCGCGGAGATGACTTGTGCAAGTTGAGC 256
Oy 81 SerGlyAlaValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGlnAla 100
Db 257 TCAGGGACCAAGGTGCGGCGACACTGTGTGGCAGAGATGACAGCACTGAGCAGCA 316
Oy 101 ProGlyLysAspThrPheThrSerLeuGlySerSerLeuAspIleThrPheAspSer 120
Db 317 CCGGGCAATGACACCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 376
Oy 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheThrAlaAlaGluAsp 140
Db 377 TACTCCAAATGAGAAGCCGTTACAGAGGCTTGTGAGCCCTTCTATGAGCGGAGATG 436
Oy 141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHis 160
Db 437 GAATGCAAGAGTGTCTGAGAGACTGAGTCCCTTGTGACCATTAATGCTCAACACTT 496
Oy 161 GlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisAspAsnLysArgThr 180
Db 497 GGGGGCTACATTTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556
Oy 181 SerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerProGlu 200
Db 557 TCAGCCCTTGTTCAGGCGCAGGTTCACAGAAATCTGGGTATCTCAGTACGCTGAG 616
Oy 201 TyrProArgProTyrProLysLeuSerSerThrThrSerLeuSerLeuGluGly 220
Db 617 TACCACAGGCATACCCCAAGCTCTCCAGTGCACCTACAGCATCCCGCGAGAGAGG 676
Oy 221 PheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGlu 240
Db 677 TTAGTGTCAATCGTGGAGCTTCTGGAGTCTTGATGTGAGAGCGCACTGAAAGCC 736
Oy 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGluHisGlyProPhe 260
Db 737 TGGCCCATATACCTCCCAAGATTCAAACAGACAGGAGGAGACCGCCCATTTTGT 796
Oy 261 LysThrLeuProHisArgIleGluThrLysSerAspThrValThrIleThrPhe 280
Db 797 AAGACGCTGCTCCCGAGATTAAGTACAGACAGGAGGAGCATCTTGTGCTGCT 856
Oy 281 AspLeuSerGlyAspHisThrGlyTyrLysIleHisThrSerThrAlaGlnPro 300
Db 857 GAGAGTGGGGAACACAGAGGTGGAATACACTACACAGACAGCAGCGCCCTG 916
Oy 301 ProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyr 320
Db 917 CCGATCCAAAGCGCGCCACTAATGCGAGCATTTGACCTGCAAGCCATATATGCT 976
Oy 321 LysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuGlnGlyHisLeu 340
Db 977 AAGGACAGGTTTCTGCTCTGCGAAGACAGGCTTCGAGCTTCTGCAAGGTCT 1036
Oy 341 LeuLysSerPheThrAlaValCysGlnLysAspLysSerThrPheAspArgPro 360
Db 1037 CTGAATATCATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1096
Oy 361 CysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGlu 380
Db 1097 TGGAGCTTATGATTTGTGCGCTCCGATGAGCTTACCAATGCGCTGTGAGCTAT 1156
Oy 381 ThrGlyProGlyValThrTyrTyrLysAlaValIleGlnTyrSerCysGluGlu 400
Db 1157 ACAGGCCCTTAAGTACTACTACAAAGCTGTGATTTAGTACAGCTGTGAGAGACT 1216
Oy 401 TyrThrMetLysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheThr 420
Db 1217 TACACATGAGCAGCAAT---GGTAATATGTGTGAGGCTGATGATTTCTGAGCAG 1273
Oy 421 SerLysGlyLysLysSerLeuProValCysGluProValCysGlyLeuSerAla 440

```

```

Db 1274 TCCAAAGAGAAAAACCTCCCGGTTTGTGAGCTTGTGTGGGCTGTCACACACT 1333
Oy 441 ThrGlyValArgIleTyrGlyGlnLysAlaLysProLysPheProGlnVal 460
Db 1334 ATAGGAGACGACATAGTTGGAGGCGACCTCGCAAGCTGTGATCTTCTTGCAAG 1393
Oy 461 LeuIleLeuGlyGlyThrThr---AlaAlaGlyAlaLeuLeuTyrAspAsnThr 479
Db 1394 TTGTGCTGGGTCAACTACAGCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTG 1453
Oy 480 ThrAlaAlaHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAspIle 499
Db 1454 ACAGCCCTCATGCTGTATGTGAGAAAGATGCGACCGCTCTCCCTGCAACATCC 1513
Oy 500 GlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaThrPheSerGluAla 519
Db 1514 GGCATCTCAAAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1573
Oy 520 HisGluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLys 539
Db 1574 CATGAAGCTACACTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1633
Oy 540 AsnLysValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAla 559
Db 1634 AACAAAGTCACAATCAACAGGAGCATCATGCTGCTGCTGCTGCTGCTGCTGCT 1693
Oy 560 SerPheMetArgThrAspAspIleGlyThrAlaSerGlyThrProGlyLeuThr 579
Db 1694 TCTTAATGAGACAGACTTCACTGCAATGTGCTGCTGCTGCTGCTGCTGCTGCT 1753
Oy 580 PheLeuAlaArgAsnLeuMetTyrValAspIleProIleValAspHisGlnLys 599
Db 1754 CTCTGCTGTAAGAAACCTAATGTTGTGACATACCATCTGCTGACCAACAAAT 1813
Oy 600 AlaAlaTyrGluLysProProTyrProArgGlySerValThrAlaAsnMetLys 619
Db 1814 GCCGTGATTAAGAAAG---CTGTATCCAGAGTAAGAGATTAAGCCGTAACAT 1870
Oy 620 GlyLeuLysSerGlyLysAspSerCysArgGlyLysSerGlyGlyAlaLeuVal 639
Db 1871 GCTTACAGACTGTGCGCAAGAGAGCTGCAAGGTGAGTGGGGGAGATTGATTTT 1930
Oy 640 LeuAspSerLeuThrGluArgTyrPheValGlyIleValSerTyrPheGlySerMet 659
Db 1931 CTGATATATGAGACACAGCATGCTTGTGCGAGATTAATTTCTGCTGCTGCT 1990
Oy 660 CysGlyGluAlaGlyTyrGlyValTyrThrLysValIleAsnTyrIleProTyr 679
Db 1991 TGTGGGCGCGCAGACACAGATGAGGCTTACACAAAGCTCATCAATATTCCT 2050
Oy 680 GluAsnIleIleSerAspPhe 686
Db 2051 GAGAACATTAATTAATTAATTTTC 2071

```

RESULT 6
AB009459 2090 bp mRNA linear ROD 13-FEB-1999
LOCUS AB009459
DEFINITION Mus musculus Masp-2 mRNA for mannose-binding lectin associated
ACCESSION AB009459
VERSION AB009459.1 GI:3928516
KEYWORDS mannose-binding lectin associated serine protease-2.
SOURCE Mus musculus (strain: BALB/c) 8 weeks liver cDNA to mRNA.
ORGANISM Mus musculus
REFERENCE
AUTHORS Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H.,
 Matsushita, M., Nonaka, M. and Fujita, T.
TITLE Two lineages of mannose-binding lectin-associated serine protease
 (MASP) in vertebrates
JOURNAL J. Immunol. 161 (9), 4924-4930 (1998)
MEDLINE 9908558

REFERENCE 2 (bases 1 to 2090)
 AUTHORS Takahashi, M., Fujita, T., Endo, Y. and Matsushita, M.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1997) Minoru Takahashi, Fukushima Medical
 College, Department of Biochemistry, 1 Hikarigaoka, Fukushima,
 Fukushima 960-1295, Japan (E-mail: minoltaecc.fmu.ac.jp,
 Tel:81-24-548-2111, Fax:81-24-548-6760)

FEATURES
 source 1. 2090
 /organism="Mus musculus"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /tissue_type="liver"
 /dev_stage="8 weeks"
 1. 2090
 /gene="Masp-2"
 33. 2090
 /gene="Masp-2"
 /codon_start=1
 /product="mannose-binding lectin associated serine
 protease-2"
 /protein_id="BAA34674.1"
 /db_xref="GI:3928517"
 /translation="MRLLFLGLMSLVATLLGSKWPEVFGRLVSPGEPEKYADHOD
 RSWTLAPPGYRLRLFTFHDLELYRCEDEYKLSGTRVATLGGESTDEQAPG
 NDIFYLGLSLKTYFSDYSENEKPTGFPAFYAEVDDECRVSLGSDVCDHYCHNYL
 GGYCSCRAGYTLHONKHKHCALSGOVFTGRSGYLSPEYPOPKLSSCTYSIRLE
 DGSYVLDPEVDFDTPHPCACPPYSLKIQDKGHSFGPCKTLEPRLETSKHTTI
 TPATDSSGHNTHKTHITSTARCPDPTAPNPNSISPVQNTVYLKRFVFCYTGTEL
 LOGSVPLKSFYAVCOQKDSMDRMPKCSIIDGPPDLPNGHVDYITGPOVYTKAVI
 QYSCETPTFTMSNGKYVCEADGFTSSGKEKLPVCEPVCGSLSTHTIGRIYGGCPA
 KPGDFPVOYLLGOTTAAGALIDHNMVLTAAHAYIEKMAASLIRNGILKRLSPH
 TYQAMEEELFIEHGYTHGAGFNDILITKKNVTNGSIIMPYCLPRKRAASIMRDFE
 TGVAGMGILFQKGLARLMPVDIPADHOKCTTYVEKLYPGRVANMLKSLTEFG
 KDSCRDSCGALVFLDNETQRMFVGGIVGSIINGAGAGQYGVTKVINTIPWENIT
 SNF"

BASE COUNT 522 a 548 c 537 g 483 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.11e-234 Length: 2090
 Score: 3062.50 Matches: 558
 Percent Similarity: 88.50% Conservative: 50
 Best Local Similarity: 81.22% Mismatches: 76
 Query Match: 81.49% Indels: 3
 DB: 10 Gaps: 3

09-874-198-2 (1-686) x AB009459 (1-2090)

1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
 33 ATAGGGCTACTCATCTCTCGGGCTCTGCTGGAGTTTGGTGGCCACACTTGGGTCA 92
 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAla 40
 93 AAGTGGCTCAACCTGATTCGGGGCTGTGCTCCCTGGCTTCCAGAGAACTATGCT 152
 41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyrArgLeuAlaGlyLeuTyr 60
 153 GACCATCAAGATGATCTGCTGACACTGACACCCCTGGCTACCCCTGGCGCTCTTAC 212
 61 PheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAspPheValValLeuSer 80
 213 TTCACCCACTTTGACCGGAACTCTTACCGGCGGAGATGACTTGTTCAGCTTACG 272
 81 SerGlyAlaValLeuAlaThrLeuGlyGlnGluSerThrAspThrGluAlaGala 100
 273 TCAGGGACCAAGGTGGCGCACACTGTGTGGCAGAGAGATACAGACACTGAGCAGCA 332
 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspThrPheArgSerAsp 120
 333 CCTGGCAATGACACTTCTACTCACTGGTCCAGCCTAAAGGTCACTTCACCTCCGAC 392

121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaAlaGluAspIleAsp 140
 393 TACTCCATGAGAACCGCTTACACAGGGTTTGAGCGCTTCTATGACAGCGAGATGGAT 452
 141 GluGlyGlnValAlaProGlyGluAlaProThrCysAspHisIscysHisAsnHisLeu 160
 453 GAATGCGAGATGCTCTCGGAGACTAGTCCCTGTGTGACCATATATGCGCACACTCTTG 512
 161 GlyIlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
 513 GCGCGCTACTATGCTCTCTCCAGCGGGCTACATTCTCCACAGAACAGCACACTGTC 572
 181 SerAlaLeuGlySerGlyGlnValPheThrGlnArgSerGlyLeuLeuSerProGlu 200
 573 TCAGCCCTTGTTCAGGCCAGCGGTTCACAGGAATCTGGATCTCAGTACCCCTGAG 632
 201 TyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGly 220
 633 TACCCGCGCATACCCCAAGACTCTCCAGCTCAGCTCAGCATCCGCGCTGGAGAGCGC 692
 221 PheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeu 240
 693 TTCAGTGTCAATCCTGGACTTCGTGGAGTCTTGATGTGGAGACGACCTGAAGCCCA 752
 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGluIscGlyProPheCysGly 260
 753 TGCCCATATACCTCCCTCAAGATTTCAACAGACAGAGGGGAACAGGCCCATTTGTGG 812
 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr 280
 813 AAGACGCTGCTCCAGGATTTGAACGTGACAGCCAGAGTGCACCTTGGCCACT 872
 281 AspGluSerCysLysPheThrGlyTyrPheLysIleHisTyrThrSerThrAlaGlnProCys 300
 873 GAGAGTGGGGGAACACACAGGCTGGAGATGACATACCAAGACAGACGCGCCCTGC 932
 301 ProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIleLeu 320
 933 CTTGATCCAAAGCGCCACCTATGCGAGACTTTCACCTGTGCAACCGCATGTGCTG 992
 321 LysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuGluGlnHisLeuPro 340
 993 AAGGACAGGTTTCTGCTCTGCTCAGACAGCGCTTGCAGGCTTGTGACAGTGTGCTCCC 1052
 341 LeuLysSerPheThrAlaValLysGlnLysAspGlySerTrpAspArgProMetProAla 360
 1053 CTGAATCATTTCACTGTGTGTCTGTCAAGAGATGATCTTGGACCGCGCGATGCGCAGAG 1112
 361 CysSerIleValAspCysGlyProProAspPheProSerGlyArgValGluTyrIle 380
 1113 TGCAGATTATGATTGTGGCCCTCCGATGACCTTACCAATGCGCATGTGGACTATATC 1172
 381 ThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerGlyGluThrPhe 400
 1173 ACAGGCCCTCAAGTACTACTCAAAAGCTGTATTCAGTACAGCTGTGAGAGACTTTC 1232
 401 TyrThrMetLysValAsnAspGlyLysTyrValCysGlnAlaAspGlyPheThrPhe 420
 1233 TACACATGACGACGAAT---GGTAAATATGTGTGTGAGGCTATGATGATTTGGACGAGC 1289
 421 SerLysGlyGluLysSerLeuProValLysGluProValCysGlyLeuSerAlaArgThr 440
 1290 TCCAAAGGAGAAACATCCCGCGGTTTGAGACCTGTTGTGGCTGTCCACACACACT 1349
 441 ThrGlyGlyArgIleTyrGlyGlnLysAlaLysProGlyLysPheProThrGlnVal 460
 1350 ATGAGGAGACGATAGTTGAGGAGCGACCTGCAAAACCTGTGTACTTTCCTTGCAAGTC 1409
 461 LeuIleGluGlyThrThr---AlaAlaGlyAlaLeuLeuTyrAspAsnThrPheValLeu 479
 1410 TTGTTGCTGGGTCAACTTACGACGAGGAGGTGCTTATATGACATATTTGGTCTTA 1469
 480 ThrAlaAlaHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAspIleArgMet 499


```

|||||.....
Db 1470 ACAGCCCTCATGCTGATATGAGAAAAGATGACACCGCTCTCCCTGACATCCGAATG 1529
Qy 500 G1YthLeuLysArgLeuSerProHisTyrThrGlnAlaIlePheSerGluAlaValPheIle 519
Db 1530 GGATCTCTAAAGGCTCTACCTCATATACATCCAAACGCGCGGCGGAAATCTTTATA 1589
Qy 520 H1G1UG1YtYrThrHisAspAlaG1YPheAspAsnAspIleAlaLeuIleLysLeuAsn 539
Db 1590 CATTGAGGCTACACTCCGCTGCTTGTGACATATATATAGCATTTGATTAACCTAAG 1649
Qy 540 AsnLysValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGlu 559
Db 1650 AACAAATCAACAATCAACGAGACATCATGCTGTGCTACCGCAAAAGAGCTGCA 1709
Qy 560 SerPheMetArgThrAspAlleGlyThrAlaSerGlyTyrPheGlyLeuThrGlnArgGly 579
Db 1710 TCTTAATGAGACAGACTTCTACTGAACTGTGGCTGGCGGCTTAAACCAGAAAGGG 1769
Qy 580 PheLeuAlaArgAsnLeuMetTyrValAspIleProIleValAspHisGlnLysCysThr 599
Db 1770 CTCTCTCTGTAAGAACTAATGTTTGTGACATACCAATTCCTGACCCCAAAATGTACC 1829
Qy 600 AlaAlaTyrG1LysProTyrProArgGlySerValThrAlaAsnMetLeuCysAla 619
Db 1830 ACCGTGATGAAAG--CTCTATCCAGAGTAAGAGTAAGCGCTACATGCTGTGCT 1886
Qy 620 G1YleuGluSerGlyGlyLysAspSerCysArgGlyLysSerGlyGlyAlaLeuValPhe 639
Db 1887 GGCTTAAAGACTGTGGCAAGACAGCTGCAGAGTACAGTGGCGGCTTAAAGTGTCTT 1946
Qy 640 LeuAspSerG1UthrGluArgTyrPheValG1YIleValSerTyrPglYSerMetLeu 659
Db 1947 CTGATATATAGACACAGCATGTTTGTGGAGAAATAGTTTCTGGGTTCCATTAAAT 2006
Qy 660 CysG1GluAlaGlyGlyTyrGlyValTyrThrLysValIleAsnTyrIleProTyrIle 679
Db 2007 TGTGGCGGCGACAGCGCATGTGGGTCTACACAAAGTCATCACTATATTCCTGGAAT 2066
Qy 680 GluAsnIleLeuSerAspPhe 686
Db 2067 GAGACATATATAGTAAATTC 2087

RESULT 7
LOCUS RNO277747 2037 bp mRNA linear ROD 05-OCT-2000
DEFINITION Rattus norvegicus partial mRNA for mannose-binding protein
associated serine protease-2 (masp-2 gene).
VERSION AJ277747.1 GI:7799288
KEYWORDS mannose-binding protein associated serine protease-2; MASP-2 gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2037)
AUTHORS Wallis R. and Dodd, R.B.
TITLE Interaction of mannose-binding protein with associated serine
proteases: effects of naturally occurring mutations
JOURNAL J. Biol. Chem. 275 (40), 30962-30969 (2000)
MEDLINE 20469449
PUBMED 10913141
REFERENCE 2 (bases 1 to 2037)
AUTHORS Wallis R.
TITLE Direct Submission
SUBMITTED (06-MAY-2000) Wallis R., Department of Biochemistry,
Physiology Institute, University of Oxford, South Parks Road,
Oxford OX1 3OU, UNITED KINGDOM
JOURNAL Location/Qualifiers
FEATURES
Source 1..2037
/organism="Rattus norvegicus"
/db_xref="taxon:10116"

```

```

/tissue_lib="liver"
1..2037
/gene="masp-2"
<1..2037
/gene="masp-2"
/function="complement activation"
/codon_start=1
/product="mannose-binding protein associated serine
protease-2"
/protein_id="CAB90832.1"
/db_xref="GI:7799289"
/translation="GLLNVATLTLGSKPEVFGRLVSPGPEKYNQHDRSWTLTA
PGFRLRLTFTHFNLSEYRCEYDVKLSTGRKVLATLGGSTDRERAPGDTFSL
GSLKVTFFHSDYSNEKPTGFARFVAEDVDCRSLGSDVPCDHCHVYLGYYCSC
RVGYLIHQHKITCSQGVYTPGRSGELSSPEYPPYRKLSACATNILEGFSITL
DVESTVDVMEHPDAQCYSLSLTIQDREKPEPCGKTLPPRIETDSNKVYITFTDES
GNHTGWKIHVTSQAPCPDPTAPPNHISVPQATVYLKDSFVFKTGELLOGSVPL
KSTAVCQKDGSMDRPIPECSIIDCGPDDPLNGHVDYITGPEVTYTKAVIYQSCET
FYTMSNGKYVCEADGFWTSKGEKSLPYCKPVGLSTHSGGRIIGGQPAKPGDPEW
OVLILGETTAAAGLIHDDKVLTAHAHVYCKTEKMSLDIRMGILKRLSPHYQAMPEA
VFTHBETHTGAGFNDIALIKLKNKTYTIRNIMPICLPKREASLMTKTPFQTVAGWG
LTQKGLFARLNLMEVDPIVDHQCATAVYTKQYPGAKVTVNMLCAGLDGGDCRCD
SGGALVFIDNETQRMFVGGIVSWSINCGSROYGYVTVKTYNIPWENIINNPF"
<1..36
/gene="masp-2"
37..2034
/gene="masp-2"
/product="mannose-binding protein associated serine
protease-2"
BASE COUNT 516 a 544 c 507 g 470 t
ORIGIN
Alignment Scores:
Pred. No.: 4.65e-234 Length: 2037
Score: 3054.50 Matches: 547
Percent Similarity: 88.51% Conservative: 54
Best local Similarity: 80.56% Mismatches: 77
Query Match: 81.28% Indels: 1
DB: 10 Gaps: 1
US-09-874-198-2 (1-686) x RNO27747 (1-2037)
Qy 8 G1YleuLysCysGlySerValAlaThrProLeuG1YProLysTyrProGluProValIlePhe 27
Db 1 GGTCTGCTTTGGAGATTGGTGCCACACTTTTGGCTCCAGTAGTGCCGTGAGCTGATATTC 60
Qy 28 G1YArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgTyrTrp 47
Db 61 GCGGCGCTGTGCCCTTGTGCCAGAGAAAGATGCAACCATGAGATTCGCTGCG 120
Qy 48 ThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuGlu 67
Db 121 ACGGTGACTCCACCCCTGCTCCGCTCGCTCTACCTTCACCCACTTCACACTTGGA 180
Qy 68 LeuSerHisLeuCysGluTyrAspPheValLysLeuSerSerGlyAlaLysValLeuAla 87
Db 181 CTCTCTTACCGCTGCGAGTGTGACTTTGTCAAGTTGACCTCAGGAGCAAGGTGCTAGCC 240
Qy 88 ThrLeuCysGlyGlnGluSerThrAspThrGluArgAlaProGlyLysAspThrPheTyr 107
Db 241 ACCCTGTGTGGCAGAGAGATACAGATACGAGCGGCGACCTGGCAATGACACTTCTAC 300
Qy 108 SerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrSerAsnGlyLysProPhe 127
Db 301 TCAGTGGGTCCACGCTAAAGGTGACCTCCACTCCGACTCCATGCAATGAGGACCATTC 360
Qy 128 ThrGlyPheGluAlaPheTyrAlaAlaGluAspIleAspGluCysGlnValAlaIleProGly 147
Db 361 ACAGAGATTAGGCTTTCTATGCAAGCGAGAGATGTGATATATGCAACATCCCTGGGA 420
Qy 148 GluAlaProThrCysAspHisHisCysHisAsnHisLeuGlyLysPheTyrCysSerCys 167
Db 421 GACTCACTCCCTTGTGACCATTAATGCGACACACTGCGGCGGCTACACTGCTCTGTCG 480

```


FEATURES	source	location/Qualifiers
REFERENCE	1. 1898	
AUTHORS	/organism="Rattus norvegicus"	
TITLE	/strain="Fisher 344"	
JOURNAL	/db_xref="taxon:10116"	
MEDLINE	/clone="prt-1"	
PUBMED	1. 1898	
REFERENCE	/gene="MASP-2"	
AUTHORS	<1. 1874	
TITLE	/gene="MASP-2"	
JOURNAL	/codon_start=3	
MEDLINE	/product="MASP-2 protein"	
PUBMED	/protein_id="CAB70973.1"	
REFERENCE	/db_xref="GI:6822243"	
AUTHORS	/translation="HFNLELSRYREYDFVKITSTKVLAITLGOESTDTERAPENDTFEYSLDSIKVTFHSDYSNEKPTGFEAFYAADVDCEKSLSDSVPCDHYCHNYLGYYCSGVGVIILHONKHHTCSALCSGVFTGRSGFLSPDEYPOPYKLSCAVNIIRGFSITLDFVSFDEVMHPEACQEPYDSLKIQTDKREYPCFGKTLPRLETDMSKVTITFTTDESNTHTGKLIHTSTAOCPDDPTAPNGHISPYOATYVILKDSVFCCKGFELGSSVPLSTPAVOCOKGDSMDRPIECSITIDGPPDLPNGHVDTGTGPEVYTYKAYOYCSCEETRYTMSNKGKYCEADGERTTSKGEKSLPVCAPVCGLSHTSGKNIIGQPAKPEDFPOVOLLIGETTAGGALIHDDWVLTAAHAYGKTEAASSLDIRNGLIKRLSLITYTQAMPPEAFVIEHGTYHAGAFNDIALILKRNKVINRNIMPICPRKEAASLMTDFVGTAAVAGWGLTKGKGLFARNLMFVDIPLVDHOKCATATYKQPYGAKTYVMLCAGIDRGKDSGCRGDSGGLVFLFNDTQRMFVGGIYMSINSIGSGSEQYGVYTKVINIPIWENIINN"	
polyA_signal	1899	
3' UTR	/gene="MASP-2"	
BASE COUNT	1875. 1898	
ORIGIN	/gene="MASP-2"	
	512 a 492 c 462 g 432 t	

		Percent Similarity:	88.62%	Conservative:	50
		Best Local Similarity:	80.61%	Mismatches:	70
		Query Match:	74.57%	Indels:	1
DB:			10	Gaps:	1
US-09-874-198-2 (1-666) x RNO18564 (1-1898)					
OY	63	HisPheAspLeuGluLeuSerHisIleu	CysGluTyrTrpAspPheValysLeuSerSergly		82
Db	3	CACCTTCAACCTGGAACTCTCTTACCGG	CTGGCAAGATGACTTGTTCACCTGACCTTCAAGG		62
OY	83	AlaIysValIleAlaThrLeu	CysGlyGlnIleuSerTrpAspTrpGlnArgAlaProgly		100
Db	63	ACCAAGGTGCTAAGCACCTGTGTGGCA	GGAAGTAAAGATACTGAACGGGCACTGGC		122
OY	103	LysAspTrpPheTyrSerLeuGlySer	LeuAspIleThrPheArgSerAspTyrSer		122
Db	123	AATGACACCTTCTACTACTGGTCCGAC	CTAAAGTCACCTTCCATCCGACTACTCC		180
OY	123	AsnGluLysProPheThrGlyPhe	GlnAlaPheTyrAlaAlaGluAspIleAspGluCys		144
Db	183	AATGAGAAGCCATTACAGAGATTGAGC	CTTCTATGACGGAGAGATGTGGAATGAC		240
OY	143	GlnValAlaProGlyGluAlaPro	ThrCysAspHisIscysHisAsnHisLeuGly		160
Db	243	AGAAATCCTCTGGGAGACTCA	AGTCCCTTGTACCAATTTTCCACAACTACTCTGGGCGC		300
OY	163	PheTyrCysSerCysArgAlaGlyTyr	ValLeuHisArgAsnHisArgThrCysSerAla		180
Db	303	TACTACTGCTCTCCGCGAGTGGCTAC	ATTCTTGCAACCAAGACATACCTGCTCAACC		360
OY	183	LeuCysSerglyGlnValPhe	ThrGlnArgSerglyGluLeuSerSerglyPro		200
Db	363	CTTTGTTCAGGCGCAGGTGTCTACT	GTGGAGTCTGGCTTCTCAGTACGCCCTGAGTATACCA		420
OY	203	ArgProTyrProLysLeuSerSergly	ThrTyrSerIleSerLeuGlnGluGlyPheSer		220
Db	423	CAGCATATCCCAAACTCTCAGAGTCC	CTPACAAATCCGCTGGAGGAAGGCTTCACT		480
OY	223	ValIleLeuAspPheValGluSer	PheAspAlaGluThrHisSerglyProGluThrLeu		240
Db	483	ATCACCCTGGACTTGTGTGAGTCTT	TATGTGGAGATGACACCTGAAAGCCACTGCCCC		540
OY	243	TyrAspPheLeuLysIleGlnThr	AspArgGluGlnHisGlyProPheCysGlyLysThr		260
Db	543	TACGACTCCCTCAACATTCAAACAG	ACAAGGGAATGAGGCCGCTTTTGTGGGAAGACG		600
OY	263	LeuProHisArgIleGlnThrLys	SerAsnThrValThrIleThrPheValThrAspGlu		280
Db	603	CTGCCCCCAGGATTGAAACTGAC	AGCAACAAGGACCACTTACCTTTACCAACGACAG		660
OY	283	SerGlyAspHisThrGlyTyrLys	IleHisTyrThrSerThrAlaGlnProCysProTyr		300
Db	663	TCAGGGAACCCACACAGCTGTGAGAT	TACACTACACAAACACACACACACACACCTGCCCTGAT		720
OY	303	PrometAlaProProAsnGlyHis	IysValSerProValGlnAlaLysTyrIleLeuLysAsp		320
Db	723	CCAAGGGGCCACCTAATGTGCAT	TTTCACCTGTGGCAAGCCAGGTATGCTCTGAAGAC		780
OY	323	SerPheSerIlePheCysGluThr	GlyTyrGluLeuLeuGlnGlyHisLeuProLeuLys		340
Db	783	AGCTTTTCTGTCTTCGAAACACT	GTGGCTTCGAGCTTTCGCAAGGTTCTGTCCCTCGAG		840
OY	343	SerPheThrAlaValCysGlnLys	AspLysSerTrpAspArgProMetProAlaCysSer		360
Db	843	TCATTCACTGCTGTGTGCAGAA	AGATGACTTGTGGACCGCGCGATATACACAGATGCAC		900
OY	363	IleValAspCysGlyProProAsp	AspAspProSerglyArgValGluTyrIleThrGly		380
Db	903	ATTATTGACTGTGGCCCTCCGAT	GAAGCTACCAATGCGCACGTTGAGACTATATACAGGC		960
OY	383	ProGlyValThrThrTyrLysAla	ValIleGlnTyrSerglyGluGluThrPheTyrThr		400

Db 248 TTCGTGAGACTCCTTGATGTGAGATGCACCCCTGAAGCCAGTCCCTACGATCCTCCT 307
 QY 247 LysIleGlnThrAspArgIuGIunHisGlyProPheCysGlyLysThrLeuProHisArg 266
 Db 308 AAGATTCAACACGACAGAGGGAATACGCCCGTTTGTGGGAAGACGCTGCCCCCAAG 367
 QY 267 IleGluThrLysSerAsnThrValThrIleThrPheValThrAspIuSerGlyAspHis 286
 Db 368 ATTGAACATGACAGACACAAAGTGACCATTTACCTTACACCGACAGATGACAGGAACAC 427
 QY 287 ThrGlyThrLysIleHisThrThrSerThrAlaGlnProCysProTyrProMetAlaPro 306
 Db 428 ACAGGCTGGAAATACACTACACAAAGCACAGCACAGCCCTGCTCATCAACGGGCA 487
 QY 307 ProAsnGlyHisValSerProValGlnAlaLysThrIleLeuLysAspSerPheSerIle 326
 Db 488 CCTAATGGTCATTTCTACCTGTGGCAAGCCACGATGTCTCTGAAGACAGCTTTTGTGC 547
 QY 327 PheCysGluThrGlyTyrGluLeuGlnGlyHisLeuProLeuLysSerPheThrAla 346
 Db 548 TTCTGCAACAGTGGCTTGACGCTTCGCAAGTTCGTCCCTGAAGTCATTTCACTGCT 607
 QY 347 ValLysGlnLysAspGlySerThrPaspArgProMetProAlaCysSerIleValAspCys 366
 Db 608 GTCTGTCAAGAAATGGATCTTGCGACCGCCGATACAGAGTGCAGCATTTATTGACTGT 667
 QY 367 GlyProPaspAspLeuProSerGlyArgValGluTyrIleThrGlyProGlyValThr 386
 Db 668 GGCCCTCCGATGACTTACCCATGGCCACGATGACTATATCACAGGCCCTGAAGTAGACC 727
 QY 387 ThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMetLysValAsn 406
 Db 728 ACTACAAGAGCTGTGATGCATACACCTGTGAAGACATTTCTACCAATAGACACCAAT 787
 QY 407 AspGlyLysTyrValCysGluAlaAspGlyPheThrThrSerSerLysGlyLysSer 426
 Db 788 ---GGTAAATATGTGTGAGGCTGATGATCTTGACAGACGTCACAAAGAGAAATCC 844
 QY 427 LeuProValCysGluProValCysGlyLeuSerAlaArgThrThrGlyGlyAlaGlyIleTyr 446
 Db 845 CTCGCCGTTTGCAGGCTGTCTGTGACCTGTCCACACACATTCACAGAGGCCGTAAAT 904
 QY 447 GlyGlyGlnLysAlaLysProGlyAspPheProTyrGlnValLeuIleLeuGlyIleThr 466
 Db 905 GGAGGACACGCTGCACAAAGCCTGGTGTCTTCTTGCAAGTCTTGTACTGGGTGAAC 964
 QY 467 ThrAlaAlaGlyAlaLeuLeuTyrAspAsnTyrValLeuThrAlaAlaHisAlaValTyr 486
 Db 965 ACAGGACAGGTGCTCTTATACATGACGACTGGTCTTAACACCGGCTCATGCTGTATAT 1024
 QY 487 GluGlnLysHisAspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSer 506
 Db 1025 GGGAAACACGAGGCGCATGTCTCTCGACATCCGATGGCATCTCTCAAAAGGCTCTCC 1084
 QY 507 ProHisTyrThrGlnAlaThrSerGlnAlaValAlaPheIleHisGluGlyTyrThrHisAsp 526
 Db 1085 CTCATTTACACTCAACGCTGGCCAGAGGCTGTCTTATCCATGAAAGGCTACACTACGGA 1144
 QY 527 AlaGlyPheAspAspAspIleAlaLeuIleLysLeuAsnAsnLysValValIleAsnSer 546
 Db 1145 GCGGTTTTTGACATGATATAGCATGATTAACTCAAGCAAAAGCTCACAAATCAACAGCA 1204
 QY 547 AsnIleThrProIleCysLeuProArgLysGlnAlaGluSerPheMetArgThrPaspAsp 566
 Db 1205 AACATATGCGCATTTGTCTACCAAGAAAGAGCTGCATCTTAATGAAACAGACTTTC 1264
 QY 567 IleGlyThrAlaSerGlyTyrGlyLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMet 586
 Db 1265 GTTGAACATGTGGCTGGCGGGGTTAACCCAGAGGGGTTCTTGCTGAACACCTAAG 1324
 QY 587 TyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProPro 606
 Db 1325 TTGTGTGACATACCAATGTGTGACACCAAAATGTGTACTGCTATACAAAGAGAGCC 1384

QY 607 TyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGlnSerGlyLys 626
 Db 1385 TACCCAGAGCAAAAGTGACTCTTTAAACATGCTGTGGCTGACCGCTGAGGGAAG 1444
 QY 627 AspSerCysArgGlyAspSerGlyAlaLeuValPheLeuAspSerGluThrGluArg 646
 Db 1445 GACACTGTAGAGGTGACAGCGGCGGCGCATTTAGTGTCTTGACAAATGCAACAGAGA 1504
 QY 647 ThrPheValGlyIleValSerTyrGlySerMetAsnCysGlyGlnAlaGlyGlnTyr 666
 Db 1505 TCGTTTGGAGGAAATAGTCTCTGGGCTCTATTACTGTGGGCGGTCAACAACATAT 1564
 QY 667 GlyValTyrThrLysValIleAsnTyrIleProTyrIleGluAsnIleIleSerAspHe 686
 Db 1565 GGGCTCTACAGCAAGTCAACGACATATTCCTGATTTGACAAACATATAATATAATTTC 1624
 RESULT 11
 AB009072 2282 bp mRNA linear VRT 14-NOV-1999
 LOCUS Xenopus laevis mRNA for mannose-binding protein-associated serine
 DEFINITION protease 2, complete cds.
 ACCESSION AB009072
 VERSION AB009072.1 GI:6407530
 KEYWORDS MASP2; mannose-binding protein-associated serine protease 2.
 SOURCE Xenopus laevis adult liver cDNA to mRNA, clone:X112.
 ORGANISM Xenopus laevis
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE
 AUTHORS Endo,Y., Takahashi,M., Nakao,M., Saiga,H., Sekine,H.,
 TITLE 1 (sites)
 JOURNAL Matsushita,M., Nonaka,M. and Fujita,T.
 MEDLINE Two lineages of mannose-binding lectin-associated serine protease
 REFERENCE (MASP) in vertebrates
 AUTHORS J. Immunol. 161 (9), 4924-4930 (1998)
 TITLE 2 (bases 1 to 2282)
 JOURNAL Endo,Y.
 TITLE Direct Submission
 FEATURES
 source location/Qualifiers
 1..2282
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="X112"
 /tissue_type="liver"
 /dev_stage="adult"
 35..2101
 /note="MASP2"
 /codon_start=1
 /product="mannose-binding protein-associated serine
 protease 2"
 /protein_id="BAA86865.1"
 /db_xref="GI:6407531"
 /translation="MTERLLASLLIMAVRPPDSTSNELTALRGIRISPGEPKRYSN
 LPTWMNIRKVPGRVRIKITYFYFNELSYLCEYDLKLSGTEVAHRCGKESTDEK
 AGDSVFPYSLDNKMTVPRSDYSNMEETGDAFYSAIDINECKETPSCDHFNNHL
 GGFYCTCRAEFTLHSDKRICOAKPDAYKRSSGRTSPDYPDAYPLFNCKYRLIVE
 EGLYTIILKEHNVNEFDVCPYDKLOLOITAKGETALLCGETIPDIIEIRSNEL
 VFEPSDGSYHKGKIOYTTKAIKCPDDILTNKGHFTPOOKRYVARDRSLTCDKGFV
 LLEDRTIPSTMTQCPNGTMDKIPKCIILYNGCPDIDNGKTYFVRAKETVLYNSV
 VQYECTGPFYTNKDKRGKRYRCGADGSECDISEKTPPVCAPDGCMKRAVARILVA
 QVASIGEPMPQVYLWNNEKGAGALLDNWILITAAHVVSYKLDSTILLIKGFLITQD
 SDYIRSPBEAVFIHEGYRQGSYDNDAIKDKVPLSESYGLGTLPRERKRYHTSDN
 EGDNHVGLVSGMGLTDRAVPSRKLEFVNITVDYSHCTLNMPRCTIFSYRMICAGF
 ETGCKDCAGDSGSLAEMDIESKKWIFGIVSMGVDCGVARQYVYIKVNYIWMIO
 KTIQSN"
 BASE COUNT 748 a 418 c 487 g 629 t
 ORIGIN

QY 674 AsnTyrIleProThrIleGluAsnIleIle 683
||||| |||||
Db 2060 AACTACTATGACTGATACGAGAAAGCAT 2089

RESULT 12
RNO18571 1112 bp mRNA linear ROD 26-JAN-2000
LOCUS Rattus norvegicus mRNA for MASP-2 protein, partial, clone prl-9.
DEFINITION Y18571
ACCESSION Y18571
VERSION Y18571.1 GI:6689098
KEYWORDS MASP-2 gene; MASP-2 protein.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1112)
AUTHORS Stover,C.M., Thiel,S., Lynch,N.J. and Schwaebler,W.J.
TITLE The rat and mouse homologues of MASP-2 and Map19, components of the
lectin activation pathway of complement
J. Immunol. 163 (12), 6848-6859 (1999)
PMID 10586086
PUBMED 2 (bases 1 to 1112)
REFERENCE Stover,C.M.
AUTHORS Direct Submission
TITLE Submitted (23-DEC-1998) C.M. Stover, Department of Microbiology
JOURNAL and, Immunology, University of Leicester, University Road,
Leicester LE1 9 HM, UK
LOCATION/Qualifiers
1. 1112
/organism="Rattus norvegicus"
/strain="Fisher 344"
/db_xref="taxon:10116"
/clone="prl-9"
1. 1112
/gene="MASP-2"
<1. 1102
/gene="MASP-2"
/codon_start=2
/product="MASP-2 protein"
/protein_id="CAB65388.1"
/db_xref="GI:6689098"
/translocation="LKDSFVFCCKTGFELLOSVPKSFYAVCKDQKSGMDRPIPECSI
IDCGPDDLPHNGHVDITGPEVYTKAVIYOGEFFYMSNGKVCADDEPMSSK
GKSLPYCKPCVCSLSTHSGRLIGGPKRDPFQVILLIGTTAGCLLHDDVLT
AAHAYGKTEAMSSLDIRKGLIKRUSLITTOAMPEAVFTHEGTHGAGDNDIALIKL
KKRVITNRNIMPICLPKREASLIMKTDFVTAGWGLTQKGLARLMLFVDIPIVDHO
KCATAYTKOYPGAKVTVMILCAGIDRGKSDCRDSCGALVELDNETORWFEVGIVS
WCSINCGSGEQQGVYTKVNYIPMIINNF"
1087. 1092
/gene="MASP-2"
1103. 1112
/gene="MASP-2"
3'UTR
polyA_signal
BASE COUNT 304 a 257 c 278 g 273 t
ORIGIN

Alignment Scores:
Pred. No.: 3,52e-121 Length: 1112
Score: 1634.50 Matches: 297
Percent Similarity: 88.56% Conservative: 28
Best Local Similarity: 80.93% Mismatches: 41
Query Match: 43.49% Indels: 1
DB: 10 Gaps: 1

US-09-874-198-2 (1-686) x RNO18571 (1-1112)

QY 320 LeuIysAspSerIlePheCysGluThrGlyTyrGluLeuGlnGlyHisLeu 339
||||| |||||
Db 2 CTGAGGACAGCTTTTCTGCTTCTGCAAGACTGGCTTCGACGTTCTGCTGTC 61

QY 340 ProLeuIysSerPheThrAlaValCysGlnIlysAspGlySerTrpAspArgProMetPro 359

Db 62 CCCCTGAAGCATTCACCTGCTGTCTGCAGAAAGATGAGATCGGACCGCGCATACCA 121
||||| |||||
QY 360 AlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGlyTyr 379
||||| |||||
Db 122 GAATGACAGCATTTATGATCTGTGGCCCTCCGATGACCTACCAATGACCGACGAGCTAT 181
||||| |||||
QY 380 IleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThr 399
||||| |||||
Db 182 ATCACAGGCCCTGAAGGACGACCTTACAAAGCTGATTCAGTACAGCTGTGAAGAGACT 241
||||| |||||
QY 400 PheTyrThrMetLysValAsnaSpGlyLysTyrValCysGlnAlaAspGlyPheTrpThr 419
||||| |||||
Db 242 TTCCTACACATGAGACGACAT---GGTAAATATGTTGTGAGCGCTGATGATTCGTGACG 298
||||| |||||
QY 420 SerSerLysGlyLysSerLeuProValCysGluProValCysGlyLeuSerAlaArg 439
||||| |||||
Db 299 AGCTCCAAAGAGAAAATCCCTCCGCTTGGACGCTGTGTGAGCTGCACACAC 358
||||| |||||
QY 440 ThrThrGlyArgGlyIleTyrGlyGlnLysAlaLysProGlyLysPheProTrpGln 459
||||| |||||
Db 359 ACTTCAGAGAGCCGTTAATTTGAGGACAGCCCTGCAAGCTGTGACTTCTCTTGCCAA 418
||||| |||||
QY 460 ValLeuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeuLeuTyrAspAsnTrpValLeu 479
||||| |||||
Db 419 GTCTGTACTGCGTGAACCTACAGCAGCAGCGTCTTATACATGACGACGCTGGCTCTTA 478
||||| |||||
QY 480 ThrAlaAlaHisAlaValTyrGluGlnIlysHisAspAlaSerAlaLeuAspIleArgMet 499
||||| |||||
Db 479 ACAGCGCTCATGCTGTATATGGGAAACAGAGCGATGCTCCCTGGACATCCGATG 538
||||| |||||
QY 500 GlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaArgProSerGluAlaValPheIle 519
||||| |||||
Db 539 GGATCTCTCAAAAGGCTCTCCCTCATTACATCAAGCTGGCCAGAGGCTGCTTTATC 598
||||| |||||
QY 520 HisGluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsn 539
||||| |||||
Db 599 CATGAAGGCTACACTCAGCAGCGCTGTTTGACAATGATATACACTGATTAACCTCAAG 658
||||| |||||
QY 540 AsnIysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluValGlu 559
||||| |||||
Db 659 AACAAAGTCAACATCAACACAAACATCATCCGATTTGTACCAAGAAAGAGCTGCA 718
||||| |||||
QY 560 SerPheMetArgThrAspAspIleGlyThrAlaSerGlyTrpIleLeuThrGlnArgGly 579
||||| |||||
Db 719 TCCTTATGAAACACACTCTGTTGACAGCTGCGCTGGGGGTTAACCCGAAGGGG 778
||||| |||||
QY 580 PheLeuAlaArgAsnLeuMetTyrValAspIleProIleValAspHisGlnLysCysThr 599
||||| |||||
Db 779 TTTCTTGTCTGAAGAACTTAATGTTGTGACATACCAATGTTGACCACAAAAATGTGCT 838
||||| |||||
QY 600 AlaAlaTyrGluLysProProTyrProArgGlySerValThrAlaAsnMetLeuCysAla 619
||||| |||||
Db 839 ACTGCGTATACAAAGCAGCCCTACCCAGCAGCAAAAGTGCCTTTAATACGCTGTGTGCT 898
||||| |||||
QY 620 GlyLeuGluSerGlyLysAspSerCysArgGlyAspSerGlyLysAlaLeuValPhe 639
||||| |||||
Db 899 GGCCTGACCGCGGTGGCAAGGACAGCTGACAGAGTGACAGCGGAGGGGATTAAGTGT 958
||||| |||||
QY 640 LeuAspSerGluThrGluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsn 659
||||| |||||
Db 959 CTAGACAAATCAACACAGAGATGTTGTGGGAGGAATGTTCCGCGGCTCTATTAC 1018
||||| |||||
QY 660 CysGlyGluAlaGlyLysIleTyrGlyValTyrThrLysValIleAspTyrIleProTrpIle 679
||||| |||||
Db 1019 TGTGGGGGCTGACAAACAGTATGGGGTCTACAGGAAAGTACAGCAATATATCTCTGGAAT 1078
||||| |||||
QY 680 GluAsnIleIleSerAspPhe 686
||||| |||||
Db 1079 GAGACATAATAATAATATTC 1099
||||| |||||
RESULT 13
MM019163

[illegible]

REFERENCE 1 (bases 1 to 3895)
 Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.
 AUTHORS Dahl, M.R., Thiel, S., Matsushita, M., Fujita, T., Willis, A.C.,
 Christensen, T., Vorum-Jensen, T. and Jensenius, J.C.
 TITLE MASP-3 and its association with distinct complexes of the
 mannan-binding lectin complement activation pathway
 JOURNAL Immunology 15 (1), 127-135 (2001)
 MEDLINE 21378425
 PUBMED 11485744
 REFERENCE 2 (bases 1 to 3895)
 AUTHORS Dahl, M.R.
 TITLE Direct Submission
 JOURNAL Submitted (03-JUL-2000), Immunology, Aarhus University, Aarhus 8000,
 Denmark

FEATURES
 source Location/Qualifiers
 1. 3895
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3q27-28"
 91..2277
 /codon_start=1
 /product="complement factor MASP-3"
 /protein_id="AAK84071.1"
 /db_xref="GI:15088517"
 /translation="MRMLLYALCFSLSKASAHVVELNNMFGQIQSPQYPSPSD
 EYTNITVYKFRILKYMFMFLNLSYLYEYVKEEDDOVLATFCGETDTQTP
 GOEVLSPGSEMSITFRSDFSNEBRTGDAMHADVDEKEREDELSCDHCHY
 TCGYVSCRFYLIHTDNTCRVECSDNLFQRTGVITSPDEPNYKPSBELLTEL
 EGGFVNILDFEDIFIEDHPEVPCPRIDYIKIVGVNVLPPFGKAPKETSIOHSVL
 ILFHSNDSNGRMLSKYRAANECPRELQPVHGKIEPQAVYFRDQVLSCDYGK
 VKLDVEMDTQIIECLKDGTWNSKIPTCKIVDCRARGLEHLIFETRNMLTYKSE
 IKYSCOEPRYKMLNNNTGILYTCSAQVMMNKVLSGLPCECQSPSLVAKRI
 IGGRAEPELFPWOLIVEDTSRVNDKMGFSGALSLSWLTAAHYRSGRDTTV
 IPVSEHNTVYGLIDHVPDKSGAVNSASRYVLHPDENQNNHDIATVLOEPYIG
 PHMVECLPRELPEEGPAPRMGLVAGMGISNPNVTYDELTISGTRLSLDYKLPV
 VHAECTSYESRSGNYSVLENMFCAGYEKGKDTCLGDSGAFVIFDLSGRWVVG
 LVSWSGPECGSKOYGYVTKVSNYVDWMVBMQGLPQSVVEPOVER"

BASE COUNT 952 a 1066 c 989 g 888 t

ORIGIN

Alignment Scores:
 Pred. No.: 1,96e-118 Length: 3895
 Score: 1609.00 Matches: 307
 Percent Similarity: 60.36% Conservative: 124
 Best Local Similarity: 43.00% Mismatches: 245
 Query Match: 42.82% Indels: 38
 Gaps: 11

874-198-2 (1-686) x AF284421 (1-3895)

QY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
 ||||| ||| ||| ||||| |||:|||||
 Db 91 ATAGAGTGGCTGCTCTCTATATGCTCTGCTTCTCCCTGCAAGGCTTACGCCAC 150
 QY 21 LysTrpPro--GluProValPheGlyArgLeuAlaSerProGlyLysPheProGlyLysTrp 39
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 151 ACCGTGAGCTAAACAATATGTTGGCCAGATCCAGTCGCCGTGTTATCCAGACTCTAT 210
 QY 40 AlaAsnAspGlnIleuArgArgTrpThrLeuThrAlaProProGlyLysArgLeuAlaGlyLeu 59
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 211 CCCAGGATTCACAGGTGACTTGGATATACATGTCGCCAGATGGGTTTCGATCAAGCTT 270
 QY 60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyLysTrpAspPheValLysLeu 79
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 271 TACTTCATGCACTTCACTTGAATCTCTACCTTTGTGAATATGACTATGTAAGGTA 330
 QY 80 SerSerGlyAlaLysValLeuAlaThrLeuGlyGlnGlnSerThrAspThrGluArg 99
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 331 GAAAGTGAGAGCAAGGTCGCGCAACCTTCTGTGGCAGGAGACCAAGACAGACAGAGCAG 390
 QY 100 AlaProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSer 119
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 391 ACTCCCGCCAGAGAGTGGTCTCTCCCTGGCTCTTCATGTCATCTTCCGGTCA 450
 QY 120 AspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaLacIleuAspIle 139
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 451 GATTTCCTCAATCAGAGAGCGTTTCAAGGCTTTGATGGCCCACTACATGGCTGTGATGTC 510
 QY 140 AspGlyCysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHis 159
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 511 GACGAGTGCACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 570
 QY 160 LeuGlyGlyPheTyrCysSerCysArgAlaGlyIleTyrValLeuHisArgLysLysArgThr 179
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 571 ATGGGCGCTACTACTGCTCTGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 630
 QY 180 CysSerAlaLeuGlySerSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerPro 199
 ||| ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 631 TGCCGAGTGGAGTGCAGTGCACAACTCTTCACTCAAAAGACAGCGGTATTCACAGCCCT 690
 QY 200 GluTyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGlu 219
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 691 GACTTCCCAACCCCTTACCCCAAGAGCTGTAATGCTGTATATCATCGAGCTGGAGAGAG 750
 QY 220 GlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGlyThr 239
 ||||| ||| ||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 751 GGTTCATGCTCAACTGCACTGAGTTGAGGACATATTTGACATTTGAGGACCATCTGAGGTG 810
 QY 240 LeuCysProTyrAspPheLeuLysIleGlnThrAspArgGlnGlnIleGlyProPheCys 259
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 811 CCTGGCCCTATGACATACATCAAGATCAAAAGTTGGTCCAAAGTTTGGGGCTTCTCT 870
 QY 260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 871 GGAGAAAGAGCCCAACACCCATCACACCCAGGACCAAGAGTCTGTGCTGTTCAT 930
 QY 280 ThrAspGlnSerGlyIleHisThrGlyIleHisThrLysIleHisThrSerThrAlaGlnPro 299
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 931 AGGACAACTCGGAGAGAACCGGGCTGAGGCTCTTACACAGGCTCAGAGAAATGAG 990
 QY 300 CysProTyrProMetAlaProProAsnGlnHisValSerProValGlnAlaLysTyrIle 319
 ||||| ||| ||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 991 TGCCGAGCTACAGGCTCTCTGTCATGGGAAATGAGCCCTCCCAACCAAGATTTC 1050
 QY 320 LeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnIleHisLeu 339
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1051 TTCAAAGACCAAGTCTGCTGACGCTGTGACACAGGCTTCAAAAGTGTGAGGATTAATGTG 1110
 QY 340 ProLeuLysSerPheThrHisValLysGlnLysAspGlySerThrAspArgProMetPro 359
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1111 GAGATGACACATTCACAGATTGAGTGTGAGAGATGGAGCTGAGTAACAAGATTCCC 1170
 QY 360 AlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluTyr 379
 ||| ||||| ||||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1171 ACCTGTAAATATGTAGACGTGAGAGCCCAAGAGACTGGAACACAGCGGTATCACCTTC 1230
 QY 380 IleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThr 399
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1231 TCTACAGAGAACACCTCCACATCAAGTGTGAGATCAAAATACCTGTCAGAGGCC 1290
 QY 400 PheTyrThrMet--LysValAsnAspGlyLysTyrValCysGluLysAspGlyPheThr 418
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1291 TATTACAGATGCTCAACATTAACACAGGATATATTAACCTGTTCTGCCAAGAGAGTCTGG 1350
 QY 419 ThrSerSerLysGlyLysLeuSerLeuProValCysGluProValCysGlyLeuSerAla 438
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1351 ATGAATTAAGTATGGGAGAAAGCCATCCACACTGCTCCAGAGTGTGTACAGCCCTCC 1410
 QY 439 ArgThrThrGlyGly-----ArgIleTyrGlyGlyGlnLysAlaLysProGlyAsp 455
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1411 CGTCCCTGCAAGCTGTGTCAAGAGATCATGGGGCGGAAATGCTGAGCTGGCCTTC 1470
 QY 456 PheProThrGlnValLeuIleLeuGlyGlyThrThrAla----- 468
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1471 TTCGCTGGAGGCTCTGATAGTGTGAGAGACACTTTCAGAGTCCAAATGACAAAGTGG 1530

QY 469 -----LalaglyalaLeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyr 486
 Db 1531 TTGGAGAGTGGGGCCCTGCTCTGCTGCTGATCTCTCAGCAGAGCTCATGCTGCTGCC 1590
 QY 487 GluGlnLysHisAspAlaSerAlaLeuAspIle---ArgmetGlyThrLeuLysArgLeu 505
 Db 1591 TCCAGCGGTAGAGACACCGGTATACCATGCTCCAAAGAGCATGTCACCGTCTACTCG 1650
 QY 506 SerProHisIleTyrThrGlnAlaTrpSerGluAla-----ValPhe 518
 Db 1651 GCGTTCATGATGTGCAGACAAATCGGGGCACTCAACAGCTCAGCTGCCAGAGTGTG 1710
 QY 519 ILethisGluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuLysLeu 538
 Db 1711 CTCACCCACCACTTC---AACATCCAAACTACACACACGATATAGCTGTGGTACAGCT 1767
 QY 539 AsnAsnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAla 558
 Db 768 CAGAGCGCTGTGGCCCTGGGAGCCCACTTATGCCCTGCTGCTGCCAAGCTTGAGCCT 1827
 QY 559 GluSerPheMetArgThrAspAspIleGlyThrAlaSerGlyThrGlyLeuThrGlnArg 578
 Db 1828 GAAGGC---CCGGCCCCCAGCATGCTGGCTGGTGGCCGGGCGCATCTCCAAATCCC 1884
 QY 579 GlyPhe-----LeuAlaArgAsnLeuMet 586
 Db 1885 AATGTGACATGATGATGATCATCAGCAGTGGCAGCAGGACTTGTGATGATGCTCTCAG 1944
 QY 587 TyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProPro 606
 Db 1945 TATGTCAAGTTACCCCTGGTGGCTCAGCGTGAAGTCAAACTACTGATGATGCTCCGCTCG 2004
 QY 607 TyrProArgIleSerValIleThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyLys 626
 Db 2005 ---GGCAATTACAGCGTCAGAGAGACATGCTGTGCTGCTGACTACGAGAGGGGCGAAA 2061
 QY 627 AspSerCysArgGlyAspSerGlyLysAlaLeuValPheLeuAspSerGluThrGlnArg 646
 Db 2062 GACACGTGCTGTGAGATACGCGTGGGCTTGTCTCATCTTTGATGACTTGAGCCAGCCG 2121
 QY 647 TrpPheValGlyGlyIleValSerTrpGlySer---MetAsnCysGlyGluAlaGlyGln 665
 Db 2122 TGGGTGTGTCAGAGCGCTGTGCTGCTGGGGGAGCTCAAGAAATGCGGCGCAGCAGCAGTGC 2181
 QY 666 TyrGlyValTyrThrLysValIleAsnTyrIleProTrpIle 679
 Db 2182 TATGAGCTTACACAAAGGTCTCCAAATTACGTGAGCTGGTG 2223
 AT 15
 AX254423 3863 bp DNA linear PART 10-OCT-2001
 LOCUS AX254423
 DEFINITION Sequence 12 from Patent WO0171004.
 ACCESSION AX254423
 VERSION AX254423.1 GI:16074167
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 3863)
 AUTHORS Yue,H., Lu,D.A., Policky,J.L., Deleage,A.M., Triboulet,C.M.,
 Khan,F.A., Au-Yang,J., Bandman,O., Lal,P., Borowsky,M.L.,
 Gandhi,A.R., Hillman,J.L., Tang,Y.T., Burford,N., Baughn,K.R.,
 Nguyen,D.B., Yao,M.G., Kalia,N.K., He,A., Lu,Y., and Patterson,C.
 JOURNAL Patent: WO 0171004-A 12-27-SEP-2001;
 Incyte Genomics, Inc. (US)
 FEATURES
 source 1..3863
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 3741323CB1"
 BASE COUNT 923 a 1063 c 991 g 886 t

ORIGIN

Alignment Scores: 2.33e-118 Length: 3863
 Pred. No.: 1608.00 Matches: 307
 Score: 60.36% Conservative: 124
 Percent Similarity: 43.00% Mismatches: 245
 Best Local Similarity: 42.79% Indels: 38
 Query Match: 6 Gaps: 11
 DB: 6
 US-09-874-198-2 (1-686) x AX254423 (1-3863)
 QY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuCysGlySerValAlaThrProLeuGlyPro 20
 Db 86 ATGAGGTGGCTGCTCTCTTATGCTCTGTCTGCTTCTCCCTGTCAAAGCTTCAGCCAC 145
 QY 21 LysTrpPro---GluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyr 39
 Db 146 ACCGTGAGCTAAACATATGTTGGCCAGATCCAGTCGCGCTGTTATCCAGACTCCTAT 205
 QY 40 AlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeu 59
 Db 206 CCCAGTATTCAGAGGTGACTTGATATCATCTGCCAATGAGGTTTCGATCCAACTT 265
 QY 60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuCysGlyLysTrpAspPheValLys 79
 Db 266 TACTTCATGCATTCACCTTGATCTCTTCTGTAATGACTATGTGAAGTA 325
 QY 80 SerSerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArg 99
 Db 326 GAACCTGAGACACGAGTGTGGCAACTCTGTGGCAGGAGACACAGACAGACAGCAG 385
 QY 100 AlaProGlyLysAspThrPheTyrSerLeuGlySerLeuAspIleThrPheArgSer 119
 Db 386 ACTCCGCGGAGGAGGTGGTCTCCCTGCTTCACTGTCACCACTTTCGGCTCA 445
 QY 120 AspTyrSerAsnGlnLysProPheThrGlyPheGlnAlaPheTyrAlaAlaGluAspIle 139
 Db 446 GATTTCCTCATGTAGAGAGCGTTTCACAGCGTTTGATGATCCCATCATAGCTGTGGATGTG 505
 QY 140 AspGlnCysGlnValAlaProGlyGluAlaProThrCysAspHisLysCysHisAsnHis 159
 Db 506 GACAGATGCAAGAGAGAGGAGAGACAGACAGCTGCTGTCAGCACTACCTCCCACTAC 565
 QY 160 LeuGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisAlaGlnLysArgThr 179
 Db 566 ATTGCGCGCTACTACTGCTCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625
 QY 180 CysSerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerPro 199
 Db 626 TGGCAGTGGAGTGCAGTGCACAACTTCTCACTCAAGAGACTGGGCTGATCCAGCCCT 685
 QY 200 GluTyrProArgProTyrTrpLysLeuSerSerCysThrTyrSerIleSerLeuGlnGlu 219
 Db 686 GACTTCCCAAAACCTTACCCCAAGAGCTGTGAATGCTGTATACCATCATGAGCTGGAGAG 745
 QY 220 GlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThr 239
 Db 746 GGTTCATGTGTCAACCGTGTGAGACATATTTTACATTGAGGACATCCGAGGTG 805
 QY 240 LeuCysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCys 259
 Db 806 CCGTGGCCCTATGATCATCATCAAGATGAGTTGGTCCAAAGTTGGTGGGCTTTTGTGT 865
 QY 260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279
 Db 866 GAGAGCAAAAGCCCAAGAACCATCAGACACACAGGACAGAGTGCCTGATCCTGTCAT 925
 QY 280 ThrAspLysSerGlyAspHisThrGlyTyrLysIleHisIleTyrThrSerThrAlaGlnPro 299
 Db 926 AGTACAACTCTGGAGAGACCGGGCTGGAGGCTCTCATACAGGCTCAGGAAATGAG 985
 QY 300 CysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIle 319

Db	966	TCGCCAGAGCTACAGACCTCCTGTCACATGAGGAAATGAGCCCTCCCAAGCAAGTATTTCC	1045
Qy	320	LeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnIleHisLeu	339
Db	1046	TTCAAGACCAAGCTGCTGCTGACGTGACACAGGCTACAAAGTGTGTGAAGATATATGTG	1105
Qy	340	ProLeuLysSerPheThrAlaValCysGlnIleLysAspGlySerTTrpAspArgProMetPro	359
Db	1106	GAGATGGACACATTTCCAGATTGATGATCTGTGCAAGGATGGAGCTGATCAAGATTTTCC	1165
Qy	360	AlaCysSerIleValAspCysGlyProProAspPleProSerGlyValArgValLeuTyr	379
Db	1166	ACCTGTAAATTTGTAGACTGTAGAGCCCCAGGAGACTGGACACGGGCTGATTCACCTTC	1225
Qy	380	IleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCysGlnLeuThr	399
Db	1226	TCTACAAGAACACACTCACACATCAAGTGTGATGATCAAAATACTGCTGTGAGAGCC	1285
Qy	400	PheTyrThrMet---LysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheTrp	418
Db	1286	TATTACAGAAAGTCTCAACAAATACACAGATATATACCTGTTCGCCCAAGAGACTGTGG	1345
Qy	419	ThrSerSerLysGlyLysSerLeuProValCysGluProValCysGlyLeuSerIleA	438
Db	1346	ATGAATTAAGTATTTGGGAGAGACTTACCCACTGCTTCCAGAGTGTGTAGCCCTCC	1405
Qy	439	ArgThrThrGlyGly-----ArgIleTyrGlyGlnIleLysAlaLysProGlyLysP	455
Db	1406	CGCTCCGTCGAAGTCTGCTGTCAGAGAGATCATTTGGGGCGCAAAATGTAGCCGGCTCC	1465
Qy	456	PheProTArgIleValIleuIleuGlyGlyThrThAla-----	468
Db	1466	TTCCGTGGCGAGGCCCTGGAATAGTGTGTGAGACACTTCCAGAGTGGCAATGACAGTGG	1525
Qy	469	-----AlaGlyAlaLeuLeuTyrAspAsnTTrpValLeuThrAlaIleHisAlaValLyr	486
Db	1526	TTTGGGAGTGGGGGCTGCTCTGCTGCTGCTGATCTTACAGCAAGCTCATGTGTGCTGC	1585
Qy	487	GlnGlnLysHisAspAlaSerAlaLeuAspIle---ArgMetGlyThrLeuLysArgLeu	505
Db	1586	TGCCAGCTAGACACACCGGTGATTCACAGTCTCCAGAGACAGATGTACCGCTTACTGTG	1645
Qy	506	SerProHisTyrThrGlnAlaTTrpSerGluAla-----ValPhe	518
Db	1646	GCGTTGACATGTATGTCGAGCAAAATGGGGGCGAGTCAACAGCTCAGCTCCCGATGCTGTG	1705
Qy	519	IleHisGlnGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeu	538
Db	1706	CTTCACCCAGACTTC---AACATTCAAAACATCAACACAGATATAGCTCTGTGTGAGCTGTG	1762
Qy	539	AsnAsnLysValIleLeaSerAsnIleThrProIleCysLeuProArgLysGluAla	558
Db	1763	CAGAGCGCTGTGGCCCGGAGACCCCAACGATTATGCCGTGCTGCCCAAGGCTTACAGCT	1822
Qy	559	GlnSerPheMetArgThrAspAspIleGlyThrAlaSerGlyTTrpGlyLeuThrGlnArg	578
Db	1823	GAAAGC---CCGGCCCCCAGCATGTGGGCTGTGGTGGCGGCTGGCATCTCCAAATCCC	1879
Qy	579	GlyPhe-----LeuAlaArgAsnLeuMet	586
Db	1880	AATGTGACATGGATGAGATCATCAGCACTGGCAGCAGGACCTTGTCAAGATGCTCTGAG	1939
Qy	587	TyrValAspIleProIleValAspHisGlnLysCysThrAlaIleTyrGluLysProPro	606
Db	1940	TATGTCAAGTATACCCGTGTGGTCCCTACAGCTGAGTGGAAACTAGCTATGAGTCCCGCTCG	1999
Qy	607	TyrProArgLysSerValThrAlaAsnMetLeuCysAlaGlyLeuGlnLysGlyIleLys	626
Db	2000	---GGCAATTACAGCTGCACGAGGAACATGTCTGTGCTGTGCTACAGAGAGGGCGCAAA	2056
Qy	627	AspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGlnArg	646

Db	2057	GACAGCTGCCTTGAGAGATACGCGTGGGGCCCTTTGTCATCTTTGATGACTTACGACGACGC	2116
Oy	647	TrpHeValGlyClyTleValSerTrrpolySer--MetAsnCysGlyLualGlyGln	665
Db	2117	TGGGGTGTGCACAGGCGCTGTCTCGGGGGGAGCCCTGAAGATCGACGACGACGCTC	2176
Oy	666	TyrGlyValTyrThrTlysValIleAsnIlyrIleProTrrIle	679
Db	2177	TATGGAGTCTACACAAAGGCTCTCCATTAACGTGGCTGGGTG	2218
RESULT 16			
AB049755	AB049755	3736 bp	mRNA
LOCUS	AB049755		Linear
DEFINITION	mus musculus masp3 mRNA for MBL-associated serine protease-3,		
ACCESSION	AB049755		
VERSION	AB049755.1	GI:16040961	
KEYWORDS	Mus musculus (strain: BALB/c) liver cDNA to mRNA.		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 Takahashi, M., Miura, S., Ishii, N., Suganuma, K., Shunichi, M., Shiro, S., Endo, Y., Matsushita, M. and Fujita, T. An essential role of MBL-associated serine protease-1 and -3 in activation of complement by lectin pathway		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 3736)		
JOURNAL	Takahashi, M., Yazawa, T., Sato, Y., Kanno, K. and Fujita, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (10-OCT-2000) Minoru Takahashi, Fukushima Medical University School of Medicine, Department of Biochemistry; 1		
JOURNAL	Hikarigaoka, Fukushima, Fukushima 960-1295, Japan		
	(E-mail: mhnolae@nu.ac.jp, Tel: 81-24-548-2111 (ex. 2232), Fax: 81-24-548-6760)		
FEATURES			
source	Location/Qualifiers		
	1..3736		
	/organism="Mus musculus"		
	/strain="BALB/c"		
	/db_xref="taxon:10090"		
	/tissue_type="liver"		
gene	1..3736		
	/gene="masp3"		
	1..53		
	/gene="masp3"		
	54..2255		
	/gene="masp3"		
5'UTR	/codon_start=1		
	/product="MBL-associated serine protease-3"		
	/protein_id="BAB69688.1"		
	/db_xref="GI:16040962"		
	/translation="MRLSPMRLILLYHALCLALPEVSAHTVELNEMFGIOLSGPYDS		
	YPSSESVTNTVTPPEGRIKLTYHMHNLSESYLCEDYVYVVEDOLAFQCRRTD		
	TECHPGRGVNVSPGETFMSVTFRSDTSVEESGNLFQRTGTVSPDNPYKPSSECS		
	YCHAYIGGYCSCRFEGYILHNDNTSCSFGSLFQRTGTVSPDNPYKPSSECS		
	YTIDLEGFMWSLQFEDIFIDEHPVPCPYDIKIKAGSKWGPFGESKSPPISTQ		
	THSVQILFRSNGSENGRSLYSPAAENECPIKLOPVYKGIPESSOAYSPKDOVLSC		
	DTGKVKLVKNGVMTPTQIECLKGASNKIPTCKIYDCAAGAGIKGLVTFSPRNLT		
	TYKKEIIRSCQPTPKMLHNTTGYTTSAGKTNTVEYLKSLPTCLPYGCOPSRALPN		
	LVKRIIGRNAEGLFEPQALIVEDTSRVPNKWFEGSALDSESWILTAAHVLRSO		
	RDNTVPIVSKEHVTVYGLHDVDRKSGAVNNSAARVTLHPDFTNIONNHDIAVOLQ		
	PVPGLAHVPICLPRPEEGPAPMLGLVAGWGISNPNTVDEILTSGTTLSDVLOK		
	VKLPVSHAECKASYESRGSNTVTEMLPCAGYEGGKDTCLDSSGAPVTFPMESGH		
	MVAAGLVSMGPEPCSGSKOYGYITKSYNTVDMLMEHNSPRAVRDLQYER"		
	2236..3736		
3'UTR	2236..3736		
	/gene="masp3"		
	3718..3723		
	/gene="masp3"		
polyA_signal			
BASE COUNT	974 a	1010 c	909 g 843 t
ORIGIN			

Db 2177 GGTCTCCACTATGTGACTGGCTA 2201
 RESULT 17
 AX156466
 LOCUS AX156466 3895 bp DNA linear PAT 22-JUN-2001
 DEFINITION Sequence 1 from Patent WO0140451.
 ACCESSION AX156466
 VERSION AX156466.1 GI:14537540
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3895)
 AUTHORS Jensenius, J.C. and Thiel, S.
 TITLE Masp 3, a complement-fixing enzyme, and uses for it
 JOURNAL Patent: WO 0140451-A 1 07-JUN-2001;
 Jensenius, Jens Christian (DK) ; Thiel, Steffen (DK)
 ES Location/Qualifiers
 1. 3895
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 91. 2277
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC42682.1"
 /db_xref="GI:14537541"
 /translation="MRRLLYALCSLSKASHVYELNNMGQIOSPEYPSYSDS
 EYWNITVPDGEFRILKLYEMHFNLESSYLEYVKETEDVLAFTCGRETDTEDTP
 GOEVLSPGSPMSITFRSDFSNEERTGDAHYMAVDVECEDELSCHYCHNY
 IGGYVSCRFEGYILHTDNRTCEKSDNLTORTGYITSPDEPNPYKSECLYTEL
 EESFNVLQEDIEDIHDHPEVPCPYDYIKIKYKPVLYPCGEGKKEPFISSOSHYL
 ILHSPNSAENEGMRLSVRAAGNECELOPAPHGKKEPBOAYFEFDVLYSCDHCYK
 YKADVENTPTECEKDKSTWSKRTPTCTIIVCRAGELEHGLIFSTNNLTXXSE
 IKYSCDEPTKMLNNNTGLYTCSAGVMMNKVLRLEPLCECGQPSRLSLVKRI
 IGRNAEPGLFPMQALIVEDTSRVNDPFGSGLLSAAILTAHVLRSSORBDYTI
 IYVSKHVTYVGLHDVRDQSGAVNSAARVYLHPDNFQNTNHDIALVQLDEPVLP
 PHMPCVCLPRELPEGPAPMLGLVAGMGSINPNVYDELISSGRTLSVLYVKLPV
 VPHAEKTSYERSGNYSVTEMNFCAGYEGGKDCDLSGSGAFVFDLISRWVVG
 LYSWGPREGCSKQYGVYTKXSNYVDWMWEDMGLPQSVYEPQVER"
 BASE COUNT 952 a 1068 c 987 g 888 t
 ORIGIN
 Alignment Scores:
 Prod. No.: 1.02e-117 Length: 3895
 Score: 1600.00 Matches: 305
 Percent Similarity: 60.228 Conservative: 125
 Local Similarity: 42.728 Mismatches: 246
 Match: 42.588 Indels: 38
 Gaps: 11
 US-09-874-198-2 (1-686) x AX156466 (1-3895)
 QY 1 MetArgLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
 Db 91 ATGAGGAGGCTGCTCTCTATATGCTGTGCTGCTCCCTGCAAGGCTTACGCCAC 150
 QY 21 LysTrpPro---GluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTrp 39
 Db 151 ACCGTGAGCTAAACAATATGTTGGCCAGATCCAGTCGCTGCTGTTATCCAGACTCTAT 210
 QY 40 AlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTrpArgLeuArgLeu 59
 Db 211 CCCAGTCATTCAGAGGCTTGGATATACACGTCGCCAGATGGGTTTCGGATCAAGCTT 270
 QY 60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAspPheValLeu 79
 Db 271 TACTTCATGCACTTCACTGGAATCCTCTACTTGTGAATATGATGCTATGGAAGCTA 330
 QY 80 SerSerGlyAlaLysValLeuAlaThrLeuGlyGlyGlnGluSerHisAspThrGluArg 99
 Db 331 GAAACGTAGAGACAGGCTGCGAACCTTCTGTGGCAGGAGGACACAGACAGAGCAG 390

QY 100 AlaProGlyLysAspThrPheThrSerLeuGlySerSerLeuAspPheThrPheArgSer 119
 Db 391 ACTCCCGGCCAGAGGTGGTCCCTCTCCCTGGCTCTTATGATGCCATCTTCCGGCTCA 450
 QY 120 AspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheThrAlaLaGluAspIle 139
 Db 451 GATTCTCCAAATGAGAGCGTTTACAGAGCTTGTATGCCACATGATGCTGTGATG 510
 QY 140 AspGluCysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHis 159
 Db 511 GACAGTCGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 570
 QY 160 LeuGlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThr 179
 Db 571 ATTTGGGGGCTACACTGCTGCTCTCCCTGGCTTACATCTCTCCACACAGACAGAGACC 630
 QY 180 CysSerAlaLeuGlySerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerPro 199
 Db 631 TGGCAGGTGAGTGCAGTGCACACCTTCTTACATCAAGGAGCTGGGTGATACACGCCCT 690
 QY 200 GluTyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGlu 219
 Db 691 GACTTCCCAACCCCTTACCCCAAGAGCTGTAATGCTGTATACATCAGCTGAGAGAG 750
 QY 220 GlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGlyThr 239
 Db 751 GGTTCATATGTCACACTGTCAGTTTGAGAGCATATTTGACATTCAGACCATCTGAGG 810
 QY 240 LeuCysProTyrArgPheLeuLysIleGlnThrAspArgGluGluHisGlyProPheCys 259
 Db 811 CCTGCCCTATGATACATCAAGATCAAAAGTTGTGCCAAAGTTTGGGGCTTCTCTGT 870
 QY 260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279
 Db 871 GGAGAAACCCCAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 930
 QY 280 ThrAspGluSerGlyLysPheHisThrGlyTyrLysIleHisTyrSerThrAlaGlnPro 299
 Db 931 AGTGACACACGCGACAGAACCGGGCTGAGGCTCTCATACAGGCTCAGAGAAATGAG 990
 QY 300 CysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIle 319
 Db 991 TGGCCAGAGCTACAGGCTCTGCTGCAATGGAAATGAGCCCTCCCAAGATTAATTC 1050
 QY 320 LeuLysAspSerPheSerIlePheCysGluThrGlyGluLeuLeuGlnGlyHisLeu 339
 Db 1051 TTCAAAGACCAAGCTCTGTCAGCTGTGACACAGGCTTACAAATGCTGGAAGGATATG 1110
 QY 340 ProLeuLysSerPheThrAlaValAlaCysGlnLysAspGlySerTrpAspArgProMetPro 359
 Db 1111 GAGATGACACATTCACAGATGAGTGTGGAAGAGATGGGACGTGAGTAACAAGATATCC 1170
 QY 360 AlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluTrp 379
 Db 1171 ACCTGTAAATATGTAGACGTGAGCCCGCAGGAGCTGGAACACGCGGTGATCACTTC 1230
 QY 380 IleThrGlyProGlyValIleThrThrTyrLysAlaValIleGlnTyrSerCysGluGluTrp 399
 Db 1231 TCTACAGGAACCAACCTCACACATACAGTGTGAGATCAAAATCTGCTGAGAGGCC 1290
 QY 400 PheTyrThrMet---LysValAsnAspGlyLysTyrValAlaCysGlnLysAspGlyPheTrp 418
 Db 1291 TATTACAAAGATGCTCAACAATACACAGGATATATATACCTGTTCTGCCAAGAGAGCTG 1350
 QY 419 ThrSerSerLysGlyLysLeuSerLeuProValCysGluProValCysGlyLeuSerAla 438
 Db 1351 ATGAATTAAGATATTGGGAGAAAGCTTACCCAGCTGCTTCCAGAGTGTGTCAGCCCTCC 1410
 QY 439 ArgThrThrGlyGly-----ArgIleTyrGlyGlyGlnLysAlaLysProGlyAsp 455
 Db 1411 CGCTCCCTCCCAAGCCTGTGTCAGAGGATCATTTGGGGCGGAATATGCTGAGCTGCGCTC 1470

Oy	456	PhePrCTPrGlnValLeuIleLeuGlyIlyThrIla-----	468
Db	1471	TTCCCGTGGCAGGCCCTGATAGTGGTGAGAGACATTCCAGAGTCCAAATGACAAAGTGG	1530
Oy	469	-----AlaGlyAlaLeuLeuTyrAspAsnTrpValLeuThrIlaAlaHisAlaValIlyr	486
Db	1531	TTTGGGAATGGGGCCCTCTCTCTGCTGCTGATCTCACAGCAGCTCAATGCTGCGCC	1590
Oy	487	GlunIlyHisAspAlaSerAlaLeuAspIle--ArgmetGlyThrLeuIlyAsArgLeu	505
Db	1591	TCCACGCTAGAGACACCAGCGGTGATTACAGTCTCCAAAGAGACATGACCGTTACTGTG	1650
Oy	506	SerProHisTyrThrGlnAlaTrpSerGluA-----ValPhe	518
Db	1651	GCGTTGCTGATGTGCGAGACAAATCCGGGGGCTGACCAACAGCTCAGTCCCGCATGCTGTG	1710
Oy	519	IleHisGluGlyTyrThrHisAspAlaGlyPheAspAsnProIleAlaLeuIleIysIleu	538
Db	1711	CTCCACCCAGACATTC--AAATCCAAACTACACACCGATATGCTCTGTGGTGCAGCTG	1767
Oy	539	AsnAsnIlyValIleValIleAsnSerAsnIleThrProIleCysLeuProArgIlyGluA	558
Db	1768	CAGAGAGCTGTGCGCCCTGGAGACCACAGCTTATGCTGTCTGCTCCCAAGCCTTGAGCCT	1827
Oy	559	GluSerPheMetAlaGlyThrAspAspIleGlyThrAlaSerGlyTyrGlyLeuThrGlnArg	578
Db	1828	GAAGCC--CCGGCCCCCACCACATGCTGGGCGCTGGTGGCGGCTGGGCGATCCCAATCCC	1884
Oy	579	GlyPhe-----LeuAlaArgAsnLeuMet	586
Db	1885	AATGTGACAGTGGATGAGATCATCATGACAGTGGCAGACAGGACCTGTCTGATGTCTGTGAG	1944
Oy	587	TyrValaAspIlePheIleValaIleAspHisGlnIlyCysThrAlaAlaTyrGluIlySerPro	606
Db	1945	TATGTCAAGTACACCGGTGTGCTGCTCACCCTGAGTGCACAACTAGCTATGAGTCCCGCTCG	2004
Oy	607	TyrProArgIlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluIlySerGlyIlyLys	626
Db	2005	--GGCAATTACAGGCTCAGCGACAGAACATGTTCGTGCTGCTACTACGAGGGCGGCAAA	2061
Oy	627	AspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGluArg	646
Db	2062	GACAGCTGCTCTGGAGATAGCGGTGGGGCTTTGTCTCATCTTGAGACTTGAGCCAGCGC	2121
Oy	647	TrpPheValGlyIlyIleValSerTrpIlySer--MetAsnGlyGluAlaGlyGln	665
Db	2122	TGGGCTGTGCAAGGCTGTCTCTCGGGGGGAGCTGTGAAGATGGGGCAGCAGCAGGCTC	2181
Oy	666	TyrGlyValTyrThrLysValIleAsnTyrIleProIlePhe	679
Db	2182	TATGAGTCTACACAAAGTCTCCAAATTACGTGAGCTGGGTG	2223
RESULT 18			
LOCUS	RNO277423	2108 bp	linear
DEFINITION	Rattus norvegicus partial mRNA for mannose-binding protein associated serine protease-1 (masp-1 gene).		
ACCESSION	AJ277423		
VERSION	AJ277423.1		
KEYWORDS	GI:7688072		
SOURCE	mannose-binding protein associated serine protease-1; masp-1 gene.		
ORGANISM	Norway rat.		
	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 2108)		
AUTHORS	Wallis, R. and Dodd, R.B.		
TITLE	Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations		
JOURNAL	J. Biol. Chem. 275 (40), 30962-30969 (2000)		
MEDLINE	20469449		
PUBMED	10913141		
REFERENCE	2 (bases 1 to 2108)		

AUTHORS
TITLE
JOURNAL

Wallis R.
Direct Submission
Submitted (14-APR-2000) Wallis R., Department of Biochemistry,
Glycobiology Institute, University of Oxford, South Parks Road,
Oxford OX1 3OU, UNITED KINGDOM

FEATURES
source
location/Qualifiers

1. .2108
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="liver"
1. .2108
/gene="masp-1"
<1. .2108
/gene="masp-1"
/function="complement activation"
/codon_start=3
/product="mannose-binding protein associated serine
protease-1"
/protein_id="CAB89695.1"
/db_xref="gi:7688073"
/translation="LSFRRLILLYHLCLLTLEVSAAHVVELNEMFGIQLSPGPDPSPYS
DSEYTNITVPEGEFRRVOLYEFMHFLPESLSECYDYVAVLENPDOLATFCGRRTDEQ
TPGEVYLSPGSFMSTVFPRSDFSNEEFTGEDATYAADVDECKEREDELSCDHCH
NIIGGVYLCDFEGFILHTDNFTCRVECSGLNFQTGSTITSIPDPNNYPKSSSESYIT
DLIEGFMYTLQDFIDIEDHEVPCEGYIKTKAGSVWGPFCGGSEPPISOSHNS
IOILFRSDNSGENRMALSYRAAGNECPKLQRPYYGKIETSOAAYSFKDOLLISCDSG
KYVLKDNEDMDTFQIECLKGAMSNK IPTCKIVDCGPAVLKHGLVTEPRNNLTTYS
SEIRYSCQOPPYKKMLHNTGTTCYCVCASGCTMNEYLKRSPLVCGLPKPSRKHSR
IFNRPRAKOGTTPWTIAMLSQNLQPFCCGSLSGSNVYLTAHCILHPLDPEELPHNS
HLSPDSDFKIIIMGMHRRSDEDDQHLLAHKIMLHPLNSTPENNDGLVLESPL
NDFFWPMICAEHPSTECTMTYVSGMGKPOFLQRPENIMEIPIPVNHTCOEAYTPAG
KKVNVPCAGEKEGKGACADGCGSMVTIKDAERDMQYLVGVVSMGECDCGKKDRGV
YSLYLPKNMDIQRTVGRN"

sig_peptide
mat_peptide
BASE COUNT 550 a 554 c 526 g 478 t
ORIGIN

Alignment Scores:

Pred. No.:	Length:	Score:
1.78e-114	2108	1555.00
Percent Similarity:	Matches:	60.64%
Best Local Similarity:	Conservative:	42.16%
Query Match:	Mismatch:	41.38%
DB:	Gaps:	10

US-09-874-198-2 (1-686) x RNO277423 (1-2108)

QY 2 ArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyProlys 21
||| ||| ||| ::||| ||| :||| |||
Db 12 AGACGGCTCTCTCTACTACCATGTCTGTGCCACCCCTG---ACGAGAGTTTCAACCATT 68
QY 22 TrpProGlu-----ProValPheGlyIyrArgLeuAlaSerProGlyPheProGlyGluTy 39
||| ||| :||| ||| :||| ||| :||| |||
Db 69 ACCGTGGAGCTAATAATGAATGTTTGCCAGATCCAGTCCGCTTATCCGAGATTCCAT 128
QY 40 AlaAsnASPngInglArgArgrThrThrLeuThrAlaProProGlyTYrArgLeuArgLeu 59
::||| ||| ||| ||| :||| ||| ||| :||| |||
Db 129 CCGAGTGCAGTCGGAGAGTGAACCTCGGAATATTACTGTCCCGGAGGGTTCCGGGTCACGTT 188
QY 60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuGlySerGlyuTYrAspPheValysLeu 79
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 TACTTCATGCAGCACTTCAACTGGAATCCTCTCACTTTGTGATATGACTATGTGAAGTA 248
QY 80 SerSerGlyIAlaIyValLeuAlaThrLeuGlySerGlyGlnGluSerThrAspThrGluArg 99
::||| ||| :||| ||| :||| ||| :||| ||| :||| ||| :||| ||| :||| ||| :||| |||
Db 249 GAACCCGAAGAAGACAGTGTGGCAACCTTCTGTGGCAGGAGACCAACCGATACCGAGAC 308
QY 100 AlaProGlyIyValSerThrPheTySerLeuGlySerSerLeuAspAlleThrPheArgSer 119

```

Db 309 ACCCGGGCCAGGAAGTGTCTTTCGCGCTTCCTTATGCTGTCACTTCCGCTCA 368
      |||||..... ||| ||||| ..: ||||| |||||
QY 120 AspTyrSerAsnGluysProPheThrGlyPheGluAlaPheYrAlaAgluAspIle 139
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 GATTTCCTCAATGAGACGATTCACAGGTTTGGACGCCCATTCATGCTGTAGATGTC 428
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 140 AspGluCysGlnValAlaProGlyGluAlaProThrCysAspHisIscysHisAsnHis 159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 429 GATGATGCGAAGAAAGGAGAGATGAGAGAGCTGCTCTGACACAGTACGCTCCCAACTAC 488
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 160 LeuGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnIysArgThr 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 489 ATCGGGGGCTACTACTGCTCCTGCTTGGCTTACATCTCCACACAGACAGACAGACC 548
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 180 CysSerAlaLeuysSerGlyValPheThrGlnArgSerGlyGluLeuSerSerPro 199
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 549 TGCAGATGGAATGACGCAATCTCTTACCCAGAGACAGGACAGCAATCACACCCCG 608
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 200 GluTyrProArgProTyrProLysLeuSerCysThrTyrSerIleSerLeuGluGlu 219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 609 GATTACCCCAACCTTACCCCAAGAGCTGATGATGTTCCATACATTGACCTGGAGAG 668
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 220 GlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThr 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 669 GCGTTCATGTTACCTGAGTTTGGAGACATTTTGACATTTGAAGACATCCCGAGGTG 728
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 LeuCysProTyrAspPheLeuysIleGlnThrAspArgGluGluHisGlyProPheCys 259
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 729 CCCTGCTCATGCTACTACATTAAAGATTAAAGCTGTTCAAAAGATATGGAGCCCTTCTGT 788
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 789 GAGAGAAATCCCCAGAACCAATCACAGCCACAGATCATAGCATTCAGATCCATTTCCG 848
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 280 ThrAspGluSerGlyAspHisThrGlyTyrPlysIleHisTyrThrSerThrAlaGlnPro 299
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 849 AGGAGCAATTCAGAGAGAACCCAGGCTGAGGCTCTCCACAGGCTGAGGAAATGAG 908
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 CysProTyrProMetAlaProProAsnGlnHisValSerProValGlnAlaLysTyrIle 319
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 909 TGCCCAAGCTACACACCCCGCTGATGGGAAATTTGAACCCGCGAGCGGTGTACTCC 968
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 320 LeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlnHisLeu 339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 969 TTCAAAGACCAAGTGTCTATACGCTGTGACAGCGGCTACAAAGTGTAAAGATACAGAG 1028
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 340 ProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspArgProMetPro 359
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 029 GTGATGAGACATTTCCAAATTGAGTGTGAAAGATGTGATGAGTAACAAGATTCCC 1088
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 AlaCysSerIleValAspCysGlyProProAspPleuProSerGlyArgValGluTyr 379
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1089 ACCTGTAATAATTGTAGACTGTGGAGTTCCAGCAGCTGCTAAACATGGGTGTAACCTTTC 1148
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 380 IleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThr 399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1149 TCCACAAAGAAACACCTCAACGCTCAATCATGATAGATGATCCGCGCCAGCAACC 1208
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 400 PheTyrThrMetLysValAsn---AspGlyLysTyrValCysGluAlaAspGlyPheTrp 418
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1209 TATTACAGATGCTTCACATATCCACAGAGGTGATATATACATGTCTCTCATGAGGACTGG 1268
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 419 ThrSerSerLysGlyLysSerLeuProValCysGluProValCysGlyLeuSer---- 437
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1269 ACAAAATGAAGTGTCAAGGAAAGCTGCCACCTCTTCCAGTGTGTGCTCCCAAG 1328
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 438 ---AlaArgThrThrGlyLysArgIleTyrGlyGlnLysAlaLysProGlyAspPhe 456
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1329 TTCTCCCGAAGACATCTCCAGAGATCTCATGAGCGCCAGCCAGAAAGGTATCACT 1388
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 457 ProTyrGlnValLeuIleLeu-----GlyGlyThrThrAlaAlaGlyAlaLeuLeu 473
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1389 CCATGATTGCCATGCTGTACAACTGAATGGGCAACCTTGTGGGGAGACCTTTTA 1448
QY 474 TyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyr----- 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1449 GGTTCCAACTGGGTTTGGACGGCTGCTCTCTACATCCACTTGTATCCAGAGAG 1508
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 487 -----GluGlnLysHisAspAlaSerAlaLeuAspIleArgMetGlyThrLeuLys 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1509 CCAATCTTACAAACATCAGATTTGCTCAGCCCTTGTGACTTCATAAAT----- 1556
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 504 ArgLeuSerProHisTyrThrGlnAlaTrpSerGlu----- 515
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1557 ATCATGGGAAGACACACGAGAGCCCGCTCAGATGATAGAGCAACAGCATCTGATTAAG 1616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 516 AlaValPheIleHisGluGlyTyrThrHisAspAlaGlyPheAspAsnIleAlaLeu 535
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1617 CACATATGCTCCACCCGCTGTAC---AACCCAGACATTTGAGACAGCATTTGGGTCTG 1673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 536 IleLysLeuAsnAsnLysValIleAsnSerAsnIleThrProIleCysLeuProArg 555
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1674 GTGAGACTGTCAAGAGCTCCAGAGCTGAGACGACTTGTGATGCTGTCTGCTGCTGAG 1733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 556 LysGluAlaGluSerPheMetArgThrAspAspIleGlyThr-----AlaSerGly 572
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1734 CACCTTCCACTGAA-----GGAACCATGCTCATGCTCAGCGCG 1772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 573 TrpGlyLeuThrGlnArgGlyPheLeuAlaArg-----AsnLeuMetYrValAsp 589
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1773 TGGGGG-----AAGAGTTCCTACAGAGCTTCCAGAGACCTGTAGGAGATTTGA 1823
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 590 IleProIleValAspHisLysSerThrAlaAlaTyrGlyLysProProTyrProArg 609
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1824 ATCCGATTTGTAACACATACATACCTCCCAAGAGCCCTAAC-----CATTTG 1871
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 Gly---SerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyLysAspSer 628
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1872 GGGAGAAAGATACCAGCAGCATGATCTGTCTGGAGAAAGAGGGGAAAGAGCGCG 1931
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 629 CysArgGlyAspSerGlyValAlaLeuValPheLeuAspSerGluThrGluArgTrpPhe 648
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1932 TGTGCTGGGACTCTGGAGGCCCATGTGTGACCAAGATGACAGAGAGACCAATGTATC 1991
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 ValGlyGlyIleValSerTrpIlePheSerMetAsnGlyGluAlaGlyIleTyrGlyVal 668
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1992 CTGCTGGCGCTGTGTGCTGCGGGGTGAA---GATTTGGGAGAAAGATGCTATGGAGTT 2048
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 669 TyrThrLysValIleAsnTyrIleProTrpIleGluAsnIle 682
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2049 TATTCTTATCTATCCCAACAAGATTTGATCCAGAGGCTC 2090
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 19
 D17525
 LOCUS D17525 4489 bp mRNA linear PRI 29-MAY-2002
 DEFINITION Human mRNA for precursor of p100 serine protease of Ra-reactive factor, complete cds.
 ACCESSION D17525
 VERSION D17525.1 GI:439712
 KEYWORDS P100 serine protease of Ra-reactive factor; CRARF; 29-kDa chain of p100; 70-kDa chain of p100; precursor of p100 serine protease of Ra-reactive factor.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 4489)
 AUTHORS Takada,F., Takayama,Y., Hatsuse,H. and Kawakami,M.
 TITLE A new member of the Cls family of complement proteins found in a bactericidal factor, Ra-reactive factor, in human serum
 JOURNAL Biochem. Biophys. Res. Commun. 196 (2), 1003-1009 (1993)
 MEDLINE 94059062
 REFERENCE 2 (bases 1 to 4489)
 AUTHORS Takada,F.
 TITLE Direct Submission

JOURNAL

Submitted (01-SEP-1993) Fumio Takada, Kitasato University School of Medicine, Pediatrics and Molecular Biology; 1-15-1 Kitasato, Sagaminara, Kanagawa 228, Japan (E-mail: ftakada@kitasato-u.ac.jp, Tel:0427-78-9115, Fax:0427-78-8441)
Submitted (01-Sep-1993) to DDBJ by:

3'UTR 2563..>4489
polyA_signal /gene="CRAR"
polyA_signal /gene="CRAR"
3397..3402
BASE COUNT 1157 a 1210 c 1098 g 1024 t

ORIGIN

Fumio Takada
Dept. of Pediatrics and Molecular Biology
Kitasato University School of Medicine
1-15-1 Kitasato, Sagaminara
Kanagawa 228
Japan
Phone: 0427-78-9115
Fax: 0427-78-8441
E-mail: ftakada@kitasato-u.ac.jp.
Location/Qualifiers

FEATURES

source

1..4489

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="3q27-28."

/clone_lib="lambda gt10 phage"

1..4489

/gene="CRAR"

<1..>4489

/standard_name="mRNA of P100 protein of Ra-reactive factor (RaRF)"

/note="human liver"

/evidence="experimental"

<1..462

/gene="CRAR"

463..2562

/gene="CRAR"

/codon_start=1

/evidence="experimental"

/product="precursor of P100 serine protease of Ra-reactive factor"

/protein_id="BAA04477.1"

/db_xref="GI:439713"

/translation="MRILLLYALCFPSLSKASAHNYELNNFGIOSPCYSPSDS
EYTNITVPDGRIRIKLYRHFHRLDESSYLCEIYVAYETDVLATFCGETTDEPT
GEEVLSRSEFSTIFRSDSNEERTGDAHYMAVDVECKEREDELSCHYCN
IGGVYSCRSFGYILHTDNRTKVECSDNLFQRTGYTSPDEPNPKSECLYEL
EGFVNWLFDEIDFIEDHPEVPCPDYIKIKVGLGPEGKAPFISTOSHVEL
ILFHSNDSNGENGRSLYRAAGNECPDELQPVHGKIEPQAYFEKDOVLVSGDYL
VKDNVEMDTFOIECLDGTSMNKIPTCKIIVDRAPELHGLTFSTNNLTYSKSE
IKYSGOEPYKMLNNNTGYTCSAGCVMMNKILGRSLPCLVYCGIPKRSKLMAIE
NCRPAOKGTPPIAMLSHLNGQPCGSGSLGSSWITVAHCLHOSLDKPTLRDLD
LSPSPFKILLGHHMLRSDENQHLGKHTTLHPKIDPMTFENDVALVLEAPVLA
FWPICLPGEPOEGAMVYVSGMKOFLORPEFTLMEIPIVDHSTCKRAVAPLKK
VTRMDICAGEKEGKDACSGDSGPMVTLNRERGVYLVGTWSGDDCKKDRGVYS
YIHHNKDIQRYTVGRN"

463..519

/gene="CRAR"

520..2559

/gene="CRAR"

/product="P100 serine protease of Ra-reactive factor"

/function="activation of C4 and C2 components of complement; bactericidal"

/standard_name="P100 protein of Ra-reactive factor (RaRF)"

/note="forms RaRF complex with mannose-binding protein; serine protease; sequence homologous to mouse P100; module homologous to C1r/C1s of complement"

520..1806

/gene="CRAR"

/product="70-kDa chain of P100"

/note="internal repeat 1; epidermal growth factor-like module; internal repeat 2; Sushi module 1; Sushi module 2"

1807..2559

/gene="CRAR"

/product="29-kDa chain of P100"

/note="protease domain"

sig_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

mat_peptide

Alignment Scores:

	pred. No.:	6.9e-114	length:	4489
Score:	1553.00		Matches:	302
Percent Similarity:	60.76%		Conservative:	130
Best Local Similarity:	42.48%		Mismatches:	233
Query Match:	41.33%		Indels:	46
DB:	9		Gaps:	14

US-09-874-198-2 (1-686) x D17525 (1-4489)

QY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20

DB 463 ATGAGTGGCTGCTCTCTATTGCTGTGCTTCCCTCCGCAAGGCTCAGCCAC 522

QY 21 LysTrpPro--GluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyTyr 39

DB 523 ACCGTGAGCTAAACATATATGTTGGCCAGATCAGTCGCTGTTATCCAGCTCTAT 582

QY 40 AlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeu 59

DB 583 CCGAGTATCAAGAGTGAAGTGAATATCACTGCCAATAGGTTTCCGATCAAGCTT 642

QY 60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGlyTyrArgPheValLysLeu 79

DB 643 TACTTCATGCACTTCAACTTGAATCCTCTACCTTTGTAATATGACTATGTGAAGTA 702

QY 80 SerSerGlyAlaLysValLeuAlaThrLeuGlyGlyGlnSerHisThrAspThrGlyArg 99

DB 703 GAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 762

QY 100 AlaProGlyLysAspThrPheThrSerLeuGlySerSerLeuAspThrPheArgSer 119

DB 763 ACCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 822

QY 120 AspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheThrAlaAlaGluAspIle 139

DB 823 GATTTCCTCAATAGAGAGGCTTTCACAGGCTTTCAGCCACATCAATAGGCTGGAGTGTG 882

QY 140 AspGluGlyGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHis 159

DB 883 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 942

QY 160 LeuGlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisAsnArgLysArgTyr 179

DB 943 ATTTGGCGGCTACTACTGCTCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1002

QY 180 CysSerAlaLeuGlySerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerPro 199

DB 1003 TGGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1062

QY 200 GlnTyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGlu 219

DB 1063 GACTTCCCAACCTTACCCCAAGAGCTGCAATGCTGTATACCAACGAGCTGGAGGAG 1122

QY 220 GlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThr 239

DB 1123 GGTTCATGCTCAACTGCGATTGAGACATATTGACATTTAGAGGACCAATCCAGAGTGTG 1182

QY 240 LeuGlyProTyrAspPheLeuLysIleGlnThrAspArgGluGluHisGlyPropheCys 259

DB 1183 CCTGCCCCATGACTACATCAAGATCAAGATGATGATGATGATGATGATGATGATGATGAT 1242

QY 260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279

DB 1242

DB 1242

DB 1242

DB 1242

DB 1242

DB 1242

DB 1242

DB 1242

DB 1242

DB 1242

Dh	1243	GGAGAGAAAGCCCGCCAGAACCCATTCACAGCCACCCAGGCCAACAGTGTCTGATCTCTGATTC	1302
Oy	280	ThraspIuSerGIyAspHisThrGIyTrpLysIleHisTyrThrSerThrAlaGlnPro	299
Dh	1303	AGTGAACAACCTCGGAGAGAACCGGGGGCTGGAGCTCTCATACAGAGGGCTCCAGAAATAG	1362
Oy	300	CysProGlyTrpMetAlaIProProAsnGlnHisValSerProValGlnAlaLysTyrIle	319
Dh	1363	TGCCAGAGCTACAGCCCTCTGTCATGGGAAATGACCCCTCCCAAGCCAAAGTATTTTC	1422
Oy	320	LeuLysAspSerPheSerIlePheCysGlnThrGIyTyrGlnLeuLencIndLysIleu	339
Dh	1423	TTCCAAAGCCAAAGTCTGCTGCACCTGTGACACAGGCTACAAAAGTCTGAGAGATATATG	1482
Oy	340	ProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTyrAspArgProMetPro	359
Dh	1483	GAGATGGACACATCTCGAGATTGAGTGTGAGAGAGATGGAGACGTGGAGTACAAAGTTGCC	1542
Oy	360	AlaCysSerIleValLAspCysGlyProProAspPLeuProSerGIyArgValGlnTrp	379
Dh	1543	ACCTGTAAATGTAGACTGTAGAGCCCGCCGAGAGACTGGACACAGGGGTGATCACTTTC	1602
Oy	380	IleThrGlyProGlyValIleThrTrpTyrLysAlaValIleGlnTyrSerCysGlnIuhTr	399
Dh	1603	TCTACAGGAACCAACCTCCACCAATACAGTCAAGATCAAAATCTCTGTAGAGAGCC	1662
Oy	400	PheTyrThrMet--LysValAsnAspGlyLysTyrValCysGlnAlaAspGlyPheTrp	418
Dh	1663	TATTACAGATGCTCACACATTAACACAGGTATATATACCTGTTCGCCCAAGAGTCTGG	1722
Oy	419	ThrSerSerLysGlyLysSerLeuProValCysGlnProValCysGlyLeuSer---	437
Dh	1723	ATGATATAAGTATTGGGAAACCTACACCACCTGCTTCCAGTGTGGGCTCCCAAG	1782
Oy	438	--AlaArgThrThrGlyGlyArgIleTyrGlyGlnLysAlaLysProGlyAspPhe	456
Dh	1783	TTTCGCCGGAAGCTGATGGCCAGAGATCTTAATGAGACGCCCAAGCAAGAGCACCT	1842
Oy	457	ProTrpGlnValLeuIleLeu-----GlyGlyThrThrAlaAlaGlyAlaLeuLeu	473
Dh	1843	CCCTGATGGCATGCTGCACACCTGAAATGGGACCCCTTGGGGAGGCTCCCTGTA	1902
Oy	474	TyrAspAsnTrpValLeuThrAlaIleHisAlaValTyrGlnGlnLysHisAspAlaSer	493
Dh	1903	GGCTCCAGCTGATCGTAGCCGCCGCGACACTGCTCCACAG-----	1944
Oy	494	AlaLeuAspIleArgMetGIyThrIleLys-----ArgLeuSerPro-----	507
Dh	1945	TCACCTCATCGGAAGATCCGACCTTACCGATTCAGACTTCTTACGCCCTTCTGCATTC	2004
Oy	508	-----HisTyrThrGlnAlaTrpSerGlnAlaValAlaPheIleHisGlnGly	522
Dh	2005	AAATCATCTCGGGCAAGCATTTGGAGGGCTCGGTACGATGAAGAAATGAAACAGATCTGGC	2064
Oy	523	TyrThrHisAspAla-----GlyPheAspAsnIleAla	534
Dh	2065	GTAACACACACCACTCTCCACCCCAAGTATGATCCCAACATTCGTAGCAATGACCTGGCT	2124
Oy	535	LeuIleLysLeuAsnAsnLysValIleIleAsnSerAsnIleThrProIleCysLeuPro	554
Dh	2125	CTGGGGAGGCGTTGGAGAGCCCAAGTGTGAATGCCCTTGATATCCCATCTGCTGCTGCT	2184
Oy	555	ArgLysGlnAlaGlnSerPheMetArgThrAspAspIleGlyThrAlaSerGIyTrpIly	574
Dh	2185	-----GAGGAGACCCCAAGCAAGAAAGAGCCATGTATGCTCATCGCCGTGGGGA	2232
Oy	575	LeuThrGlnArgGlyPheLeuAlaArg-----AsnLeuMetTyrValAspIlePro	591
Dh	2233	-----AAGCAGTCTCTTCCAAAGGTTTCCAGAGACCTCATGATGAGATTGAATTCGG	2283
Oy	592	IleValAspHisGlnLysCysThrAlaIleTyrGlnLysProProTyrTrpArgGlySer	611
Dh	2284	ATTGTGTACACAGCACCTCTCCGAAAGGCTGTAT-----GCCCCCTCAAGAG-----AAA	2334

```

QY 612 VAlThrrAlAaSmLeLcCysAlaGlyLeuLcUsErGlyGlyAspSerCyArGly 631
Db 2335 GtGACCAGGAGCATGATCTGCTGGGAGAGAGAGAGGAGGAGGAGCCTGTTTCGGCT 2394
QY 632 AsPserGlyGlyAlaLeuValPheLeuAspSerGluTrGluArGTrPpPheValGly 651
Db 2395 GACCTGGAGGCCCCCATGTGTACCTGAAATGAGAGAAAGGCCAGTGTACTGTGGTGGC 2454
QY 652 lLeValSerTrpGlySerMetAsnCysGlyGlyAlaGlyGlnTrpGlyValTrpTrLys 671
Db 2455 ACTGTGCTCTGGGGTGAN---GACTGTGGAGAAAGAGCACCGCTAGGATACTCTTAC 2511
QY 672 VAlleAsnTrpLleProTrpLleGluAsnLe 682
Db 2512 ATCCACCAACACAGGACTGATCCAGAGGGTC 2544

RESULT 20
LOCUS MUSCRARF
DEFINITION Mouse mRNA for P100 serine protease of Ra-reactive factor (RaRF),
complete cds.
ACCESSION D16492
VERSION D16492.1 GI:403200
KEYWORDS 29-kDa chain; 70-kDa chain; P100 serine protease; Ra-reactive
factor; RaRF.
SOURCE Mus musculus (sub-species:domesticus, strain:BALB/c) liver cDNA to
mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Takahashi,A., Takayama,Y., Hattuse,H. and Kawakami,M.
TITLE Presence of a serine protease in the complement-activating
component of the complement-dependent bactericidal factor, RaRF, in
mouse serum
JOURNAL Biochem. Biophys. Res. Commun. 190 (3), 681-687 (1993)
MEDLINE 93176166
REFERENCE 2 (bases 1 to 5135)
AUTHORS Takayama,Y., Takada,F., Takahashi,A. and Kawakami,M.
TITLE A 100-kDa protein in the C4-activating component of Ra-reactive
factor is a new serine protease having module organization similar
to Clr and C1s
JOURNAL J. Immunol. 152 (5), 2308-2316 (1994)
MEDLINE 94179811
REFERENCE 3 (bases 1 to 5135)
AUTHORS Takayama,Y.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1993) Yoshinaga Takayama, Kitasato University,
Dept. of Mol. Biol., School of Medicine; 1-15-1 Kitasato,
Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9115,
Fax:0427-78-8441)
COMMENT On Sep 25, 1993 this sequence version replaced gi:303703.
Submitted (19-JUN-1993) to DDBJ by:
Yoshinaga Takayama
Dept. of Molecular Biology
School of Medicine
Kitasato University
1-15-1 Kitasato, Sagamiara
Kanagawa 228
Japan
Phone: 0427-78-9115
Email: ytakayam@iguts.nig.ac.jp
Fax: 0427-78-8441.

FEATURES
Source
1..5135
/organism="Mus musculus"
/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/chromosome="16"
/map="16B2-B3."
/tissue_type="liver"

```



```

Db 1407 TCCAGATGTTCAATGAGCCGCCAGCAAGATACCATGATGATGCCATGCTG 1466
QY 463 Leu-----GlyGlyThrThrAlaAlaGlyAlaLeuLeuTyrrspantprvalLeu 479
Db 1467 TCACACCTGACGAGAACACCTTCTGTGGGAGAGCTTTAGCTTCAACCTGGCTTTT 1526
QY 480 ThrAlaAlaHisAlaValTyrrGlnGlnHisAspAlaSerAlaLeuAspIleArgMet 499
Db 1527 ACAGCTGCTACATGCTCCACAG-----TCATTGTATCCAGAAGAA 1568
QY 500 GlyThrLeuLysArg-----LeuSerPro----- 507
Db 1569 CCAACCTTACACAGCTCATACCTGCTCAGCCCTTTCATTCAAAATTATCATGGCAAG 1628
QY 508 HisTyrThrGlnAlaIleArgSerGlu-----AlaValIleIle 519
Db 1629 CACTGGAGAGCGGCTCAGACAGACGACGACCTGCATGTAAAGCCGACCGCTC 1688
QY 520 HisGlnGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsn 539
Db 1689 CACCCACCTGAC---AACCCACGACGCTTGAAGACGACCTTGGTCTGGTGAAGTGTCA 1745
QY 540 AsnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGlnAlaGlu 559
Db 1746 GAGAGCCGAGAGCTGACGACGCTTGTGATGCTGTCTGCTGCAGCAGCCTTCACAT 1805
QY 560 SerPheMetArgThrAspIleGlyThr-----AlaSerGlyTyrPglyLeuThr 576
Db 1806 GAA-----GGAACCATGTCATCTCATCTGCTGCTGGGG----- 1838
QY 577 GlnArgGlyPheLeuAlaArg-----AsnLeuMetTyrValAspIleProIleVal 593
Db 1839 ---AACGAGTCTTACAGAGCTTCCAGAGACCTGATGAGATGTGAATCCCAATTGTA 1895
QY 594 AspHisGlnLysCysThrAlaAlaTyrGlnLysProProTyrProArgLysSerValThr 613
Db 1896 AACTGTGACACTGCGACGAGCGCCTAT-----ACCCATTGAAGAAG---AAAGTGACC 1946
QY 614 AlaAsnMetLeuLysAlaGlyLeuGlnSerGlyGlyLysAspSerCysArgLysAspSer 633
Db 1947 AAGGACATGATGCTGTCCCGAGAAAGGAGGGGGAAGATGCTGTGCTGCTGACTCT 2006
QY 634 GlyGlyAlaLeuValPheLeuAspSerGluThrGluArgTyrPheValGlyGlyIleVal 653
Db 2007 GAGGCGCTATGATGACCAAGATGACAGAGACCAATGATACCTGGTGGCGGTG 2066
QY 654 SerTyrPglySerMetAsnCysGlyGlnAlaGlyGlnTyrGlyValTyrThrLysValIle 673
Db 2067 TCCTGGGGTGAA---GATTGCGGGAAGAAAGATGCGATGACTCTATCTTACATCTAT 2123
QY 674 AsnTyrIleProTyrIleGluAsnIle 682
Db 2124 CCCACAGACGCTGATCCAGAGATC 2150

RESULT 21
LOCUS HUMASP 2787 bp mRNA linear PRI 29-MAY-2002
DEFINITION Human mRNA for MASP (mannose-binding protein associated serine
ACCESSION D28593
VERSION D28593.1 GI:790963
KEYWORDS C1s like serine protease; MASP; MBP associated serine protease;
SOURCE mannose-binding protein associated serine protease.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2787)
AUTHORS Sato,T., Endo,Y., Matsushita,M. and Fujita,T.
TITLE Molecular characterization of a novel serine protease involved in
activation of the complement system by mannose-binding protein

```

```

JOURNAL Int. Immunol. 6 (4), 665-669 (1994)
MEDLINE 94289349
REFERENCE 2 (bases 1 to 2787)
AUTHORS Sato,T.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1994) Tetsuo Sato, Fukushima Medical College,
Department of Biochemistry, 1 Hikarigaoka, Fukushima, Fukushima
960-12, Japan (E-mail:GBD02047@niftyserve.or.jp,
Tel:0245-48-2111(x.2232), Fax:0245-48-6760)
COMMENT On May 2, 1995 this sequence version replaced gi:471127.
Submitted (22-FEB-1994) to DDBJ by:
Tetsuo Sato
Department of Biochemistry
Fukushima Medical College
1 Hikarigaoka, Fukushima
Fukushima 960-12
Japan
Phone: 0245-48-2111 x2232
Email: GBD02047@niftyserve.or.jp
Fax: 0245-48-6760.

FEATURES
source location/Qualifiers.
1..2787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3q27-28."
/sex="male"
/tissue_type="liver"
/clone_lib="lambda gt10"
/dev_stage="26 week old fetus"
91..2190
/function="C1s like Serine protease"
/standard_name="MBP associated Serine Protease MASP"
/codon_start=1
/product="MASP"
/protein_id="BAA05928.1"
/db_xref="GI:471128"
/translation="MRWLLLYALCFSLSKASAHVVELNNMFGIOLSPGYPSPSOS
EVTWNITVDPGFRILKLYEMHLENLSESYLCEYDYVVEEDQVLAIFPGRETDTPTP
GOEVLVSPGSRMSITFRSDFSNBEERFTGDATYMAVDDECKEREDELSCHYKHYNY
ICGVYSCRFEGILHTDNKTCVBCSDNLFORTGVITSPDPNYPKSSSECLVYTEL
EEGFVNQLQFEDIFDIQDHPVEPCPDYIKIKVGRKVLGPPGGEKAPRISTQSHSVL
ILFHSDNSAENGRMSLYRAAGNECPDELQPVHCKLEPSQAKYFFKDDVLVSCDGYK
VLKDNDMDTFOIECLKDGTMSNKIPTCKIYDCRAGDEHGLITFSTRNNLTYYKSE
IKYSCQEPYKMLNNNTGICYSCAGVMNNKYIGRSLPTCLPYGGLPFRSRYLMARIF
NCRPRQKGTTPWITANLSHNGOPPCGSLSSNLYVTRACHLHOSLDPCDPTLRSDI
LSPDFKTIILGKHMRLKSDENQHLGVTTLHPQYDNPTEENDVALVELLESPLNA
FVMPYICLPEGPQDEGAMNIVSGMGKQFLQRPETLMELIPIVDHSTOKAAYAPLKK
VTRDMICAGEKEGKGDACAGDSGGPMVTLNREBGQMYLVGTVMGDDCGKDRYGVYS
YIHNKMDMTQRYTVGN"
91..147
sig_peptide
mat_peptide
148..2187
/product="MASP"
/standard_name="MASP"
2787
polya_site
BASE COUNT 713 a 773 c 689 g 612 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

Db      151 ACCGTGGACCTAAACATATGTTTGGCCAGATCCAGATCCGCTGTATCCAGACTCTTAT 210
Oy      40 AlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyIleArgLeuArgLeu 59
Db      211 CCCAGATTCAGATGAGTGAAGTATACAGTCTCCAGATGCTTGGATCAAGCTT 270
Oy      60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAspPheValIleu 79
Db      271 TACTTCATCCTCACTCAACTTGGAAATCCTCTCACTTGTGGAATATACATATGGAAGTA 330
Oy      80 SerSerGlyAlaIleValIleuAlaThrLeuGlyGlnGluSerThrAspThrGluArg 99
Db      331 GAAACTGAGACGACGAGCTGGCAACTCTGTGGCAGAGAGACACAGACAGAGCAG 390
Oy      100 AlaProGlyIleAspThrPheThrSerLeuGlySerSerLeuAspIleThrPheArgSer 119
Db      391 ACTCCCGCCAGAGAGGTCTCTCTCCCTGCTCTTCACTGCTCCACTTTCCTGCTCA 450
Oy      120 AspTyrSerAsnGlnIleuSerProPheThrGlyPheGluAlaPheTyrAlaIleGluAspIle 139
Db      451 GATTTCCTCAATGAGAGAGCGTTTCAACAGCCTTGGATGCCACTACATGGCTGTGGATGTG 510
Oy      140 AspGluCysGlnValAlaProGlyIleAlaProThrCysAspHisHisCysHisAsnHis 159
Db      511 GACGAGTGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
Oy      160 LeuGlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnIleArgThr 179
Db      571 ATTTGGGGTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630
Oy      180 CysSerAlaLeuGlySerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerPro 199
Db      631 TCCGCGATGGAGTGAAGTGAACAACCTTCACTCAACAAGAGAGAGAGAGAGAGAGAGAG 690
Oy      200 GluTyrProArgProTyrProIleuSerSerCysThrTyrSerIleSerLeuGluGlu 219
Db      691 GACTTCCCAACCCCTTACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 750
Oy      220 GlyPheSerValIleLeuAspPheValIleuSerPheAspValIleuThrHisProGluThr 239
Db      751 GGTTCATATGCTCACTGCACTTGTGAGACATATTTGACATTTGACAGACATCTTGAGGTG 810
Oy      240 LeuGlyProTyrAspPheLeuIleGlnThrAspArgGluGlnIleGlyProPheCys 259
Db      811 CCTGCTCCCTATGACTACATCAAGATCAAGTGTGCTCAAAAGTTTGGGGCTTCTCTGT 870
Oy      260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279
Db      871 GGAGAGAAAGGCCCCAGAACCCATCAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 930
Oy      280 ThrAspGluSerGlyAspHisThrGlyTyrLysIleHisTyrThrSerThrAlaGlnPro 299
Db      931 AGTGACAACTCCGACAGAGAACCGGGCTGGAGGCTCTCATACAGGGCTCCAGAGAAATGAG 990
Oy      300 CysProTyrProMetAlaProProAsnGlnHisValSerProValGlnAlaIleTyrIle 319
Db      991 TGGCCAGAGCTACAGCCCTCTGTCATGGGAAATGAGCCCTCCCAACAGCAAGTATTTTC 1050
Oy      320 LeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlyHisLeu 339
Db      1051 TTCAAAGACCAAGTCTCTCAGCTGTGACACAGAGCTTACAAAGTCTGGAAGGTAATGTG 1110
Oy      340 ProLeuLysSerPheThrAlaValCysGlnLysAspGlySerThrAspArgProMetPro 359
Db      1111 GAGATGGAACATTCACAGATTGAGTGTCTGAAGAGATGGAGCGTGGAGTACAAAGATTTCC 1170
Oy      360 AlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyIleValGluTyr 379
Db      1171 ACCTGTAAATATGTAGACTGTAGAGCCCAAGAGAGCTGGAACAGAGGCTGATCACCTTC 1230
Oy      380 IleThrGlyProGlyValThrThrTyrLysAlaValIleGlnIleTyrSerCysGluGluThr 399

```

```

Db      1231 TCTACAGAGAACACCTCACCATPACAGTCTGAGATCAATACTCTGTACAGAGCCC 1290
Oy      400 PheTyrThrMet---LysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheTrp 418
Db      1291 TATTCACAGATGCTCACACATATACACAGGATATATACCTGTCTTCCCAAGAGCTTGG 1350
Oy      419 ThrSerSerLysGlyGluLysSerLeuProValCysGluProValCysGlyLeuSer--- 437
Db      1351 ATGAATAAAGTATTTGGGAGAAAGCCATCCAGCTGCTTCCAGTGTGTGGGCTCCCAAG 1410
Oy      438 ---AlaArgThrThrGlyLysArgIleTyrGlyGlnLysAlaLysProGlyAspPhe 456
Db      1411 TTCTCCCGAAGCTGATGCGCAGAGATCTTAATGACAGCCCAAGCCAGAAAGCAGCACT 1470
Oy      457 ProTrpGlnValLeuIleLeu-----GlyGlyThrThrAlaIleAlaLeuLeu 473
Db      1471 CCTGGATTGGCCATGCTGTACACCTGAAATGGGACCCCTTCTGCGAGAGCTCCCTTCA 1530
Oy      474 TyrAspAsnTrpValIleuThrAlaAlaHisAlaValTyrGluGlnLysHisAspIleSer 493
Db      1531 GGCTCAGCTGGATGCTGACCGCCGACACCTGCTCCACAG----- 1572
Oy      494 AlaLeuAspIleArgMetGlyThrLeuLys-----ArgLeuSerPro----- 507
Db      1573 TCATCTGATCCGGAGATCCGACCCCTACCTGATTCAGACTTGTCTGACCTTCTGACTTC 1632
Oy      508 -----HisTyrThrGlnAlaTrpSerGlnAlaValAlaPheIleHisGluGly 522
Db      1633 AAATCATCTCTGGGCAAGATTTGAGGCTCCGCTCAGATGAATGAACAGACATCTCGGC 1692
Oy      523 TyrThrHisAspAla-----GlyPheAspAsnAspIleAla 534
Db      1693 GTCAAAACACACCACTCTCACACCCCGATATGATCCCAACACATTCAGATATGACGTGGCT 1752
Oy      535 LeuIleLysLeuAsnAsnLysValValIleAsnSerAsnIleThrProIleCysLeuPro 554
Db      1753 CTGGTGAGCTGTGGAGAGAGCCAGCTGAAATGCTTGTGATGCCATCTGTCTGCTCT 1812
Oy      555 ArgLysGlnAlaGluSerPheMetArgThrAspAspIleGlyThrIleAspGlyTyrTrpGly 574
Db      1813 -----GAGGAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Oy      575 LeuThrGlnArgGlyPheLeuAlaArg-----AsnLeuMetTyrValAspIlePro 591
Db      1861 -----AAGAGTTCCTTGCAAAGTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1911
Oy      592 IleValAspHisGlnLysCysThrAlaAlaTyrGluLysProProTyrProArgGlySer 611
Db      1912 ATTTGTTGACCACAGCAGCTGCGCAGAGAGGCTTAT-----GCCCGCTGAAAGAG--AAA 1962
Oy      612 ValThrAlaAsnMetLeuGlyValGluLeuGluSerGlyGlyLysAspSerCysArgGly 631
Db      1963 GTGACAGGAGGACATGATGTGTGTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
Oy      632 AspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGluArgTrpPheValGlyGly 651
Db      2023 GACTCTGGAGGCCCATGTGATGACCCCTGAATPAGAGAAAGAGCCAGTGTACCTGTGGGGC 2082
Oy      652 IleValSerTrpGlySerMetLysCysGlyGluAlaGlyGlnIleTyrGlyValTyrThrLys 671
Db      2083 ACTGTCTCTGTGGGTGAT---GACTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139
Oy      672 ValIleAsnTyrIleProTrpIleGluAsnIle 682
Db      2140 ATCCACACACACAGAGACTGATCCAGAGGCT 2172

RESULT 22
RNO18565      1027 bp      mRNA      linear      ROD 27-JAN-2000
LOCUS      RNO18565      Ratius Norvegicus mRNA for MASP-2 protein, partial, clone p1-2.
DEFINITION      Y18565.1 GI:6689086
ACCESSION      Y18565.1
VERSION      MASP-2 gene; MASP-2 protein.
KEYWORDS

```

SOURCE ORGANISM	
<i>Rattus norvegicus</i> .	
<i>Rattus norvegicus</i>	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
<i>Rattus</i> .	
REFERENCE	1 (bases 1 to 1027)
AUTHORS	Stover,C.M., Thiel,S., Lynch,N.J. and Schwaebler,W.J.
TITLE	The rat and mouse homologues of MASP-2 and MAP19, components of the lectin activation pathway of complement
JOURNAL	J. Immunol. 163 (12), 6848-6859 (1999)
MEDLINE	20054576
PUBMED	10586086
REFERENCE	2 (bases 1 to 1027)
AUTHORS	Stover,C.M.
TITLE	Direct Submission
JOURNAL	Submitted (23-DEC-1998) C.M. Stover, Department of Microbiology and Immunology, University of Leicester, University Road, Leicester LE1 9 HN, UK

TITLE Direct Submission JOURNAL Submitted (21-NOV-1997) Yuichi Endo, Fukushima Medical College, Dept. of Biochemistry, 1-Hikarigaoka, Fukushima, Fukushima 960-12, Japan (E-mail: yendoecc.fmu.ac.jp, Tel: 0245-48-2111 (ex. 2233), Fax: 0245-48-6760)

FEATURES
source Location/Qualifiers

1. 2407
/organism="Triakis scyllium"
/db_xref="taxon:30494"
/tissue_type="liver"
/dev_stage="adult"
51.. 2210
/note="MASP"
/codon_start=1
/product="mannose-binding lectin-associated serine
protease"
/protein_id="BA86867.1"
/db_xref="GI:6407545"

CDS

BASE COUNT 652 a 553 c 610 g 592 t
ORIGIN
Alignment Scores:
Pred. No.: 8,186-111 Length: 2407
Score: 1510.00 Matches: 291
Percent Similarity: 58.08% Conservative: 126
Best Local Similarity: 40.53% Mismatches: 245
Query Match: 40.18% Indels: 56
DB: 5 Gaps: 14

US-09-874-198-2 (1-686) x AB009074 (1-2407)
13 SerValAlaThrProLeuGlyProLysTrpProGlu-----ProValPheGlyArgLeu 30
78 AGCTTCTTTACCCCTGCTGAGCAATGTGACTGAACTGACTGGAGTTTGGGAGATT 137
31 AlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThr 50
138 CAGTCTCCCAATTATCCGGAAGATGCCAAGATTGACCTTAACCTGGAATATATACA 157
51 AlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuGluLeuSerHis 70
198 GTACCGCGGTGATTGCAATCAAGCTTACTCATGCACTTTCACCTGGAGCCCTCTAT 257
71 LeuGlyGluTyrAspPheValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuCys 90
258 CTCTGTGAATACACTACGTAAGAGTTGTCTGATGAGAGATTGCTGGAAACATCTCTG 317
91 GlyGlnGluSerThrAspThrGluArgAlaProGlyLysAspThrPheTyrSerLeuGly 110
318 GCGAGAGAGAGACTGACTGAGCAATCCAGGCCGAGATTGTCCTGACTATACAGC 377
111 SerSerLeuAspThrPheArgSerAspTyrSerAsnGlyLysProPheThrGlyPhe 130
378 AATACATGCTGCTGCAATTGAAGACTGACTTTCACATAGAGAGCGTTTCACCGGCTTC 437
131 GluAlaPheTyrAlaAlaGluAspLeuAspGluCysGlnValAlaLapProGlyGluAlaPro 150
438 GAGGCTCATATATCTGCGCGTCGATATCGATGATGTGTGGAGAACTAGTAGAGCGTTA 497
151 ThrCysAspHisThrCysHisAsnHisLeuGlyGlyPheTyrCysSerCysArgAlaGly 170

498 GCTGTGACCATTTCTGTACCAATTATACATCGGTGATATTACGTCCTGACATTTAGC 557
171 TyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSerGlyGlnValPheThr 190
558 TACATTCGACACACATATATCGACCTGTAAGATTGAGTGAATGACATGATTTACAT 617
191 GluArgSerGlyLeuLeuSerSerProGluTyrProArgProArgProLysLeuSerSer 210
618 GAAAGTCAGAGATTATACAGAACTCCAGATTCACATTCATATCCAAAGCTCCGAC 677
211 CysThrTyrSerIleSerLeuGluGluGlyPheSerValIleLeuAspPheValGluSer 230
678 TGTCTGTACCGCATTTGACACTTGAGGAAGTTTGGCCATCTCGTTGGAGTTTGGACATC 737
231 PheAspValGluThrHisProGluThrLeuCysProTyrAspPheLeuLysIleGlnThr 250
738 TTTGATTTTACAGACACACCCAGATGTTCCCTGTCGATACATCTGTAAGATTAAAGCT 797
251 AspArgGluGluHisGlyProPheCysGlyLysThrLeuProHisArgIleGluThrLys 270
798 GGATTCGATGCGCTTGTGCTTCTGTGGGATAAAGTCCAGAGAAAGATGATACAAAC 857
271 SerAsnThrValThrIleThrPheValThrAspGluSerGlyAspHisThrGlyTyrLys 290
858 AGCAGTAGTGTGATGATTTCTTCCATATAGCGGGCGCTCAGAGGATTAACAAAGCTGGA 917
291 IleHisTyrThrSerThrAlaGlnProCysProTyrProMetAlaProProAsnGlyHis 310
918 ATGTGATTTCTCTCAGTTGAGTGAACCTCTTGTCTTCATCTCCATCTCTCCGCAATGCC 977
311 ValSerProValGlnAlaLysTyrIleLeuLysAspSerPheSerIlePheCysGluThr 330
978 ATTGACACTGTCACAGACCAATATTTCATTCAAAGATCACTCCAGATCAGATCAACACT 1037
331 GlyTyrGluLeuLeuGlnGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLys 350
1038 GGAATTAATATTGTAAGCGTGTGATGAGTGGAGGCTATCAATATCAAAATGCCAGAA 1097
351 AspGlySerThrAspArgProMetProAlaCysSerIleValAspCysGlyProProAsp 370
1098 TCCGGAAGCTGGAGTAGACACATTTCCACATGTAAATGATTGCTGGCGGACCGCA 1157
371 AspLeuProSerGlyArgValGluTyrIleThrGlyProIleValThrThrValAla 390
1158 CCCCTGAGAGATGAATCAATCAATTCATTCGAAAGAAACCTCACACTGTTCAAGAGC 1217
391 ValIleGlnTyrSerCysGluGluThrPheTyrThrMetLysValAsn---AspGlyLys 409
1218 CACATGAAATACCTGCTGCTGCTTACATTTACCTGAGCCGAGCTGATATACACT 1277
410 TyrValCysGluAlaAspGlyPheThrPheThrSerSerLysGlyLysSerLeuProVal 429
1278 TACACCTGTGTGAGATGATTTTGGTCAATGATGACCTGGGACACCGACCTCCCTACC 1337
430 CysGluProValCysGlyLeuSerAlaArgThrThr-----GlyGlyArgIleTyr 446
1338 TGTACGCACTGTGTGAGAGCCCTGCTGACCCCTGCCACACACAAAGCGGATTAAT 1397
447 GlyGlyGlnLysAlaLysProGlyAspPheProTyrProIleValLeuIleLeuGlyThr 466
1398 GAGAGGGGAGCTGAGCCGACGCTTCTCCCTGCGAGTTATGATGAGTGGAGAGAT 1457
467 ThrAla-----AlaGlyAlaLeuLeuTyrAspAsnTyr 477
1458 GTGAGCGAGTCCCAAGAGCAAGTGTGAGGAGTGTGAGGAGTGTGCTTCGACAGATG 1517
478 ValLeuThrAlaIleHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeu----- 495
1518 GTACTGACACAGCCACCGCTGTTGCCCTACAAAGAGGAGATACATCACTGTGTTCCA 1577
496 -----AspIleArgMetGlyThrLeuLysArg 504

Db 1578 AGTGAATATGCTACTATCTATGCTGGCTCCATGACGCTCCGACAG-----AAGGAA 1628
Qy 505 LeuSerProHisThrGlnAlaIleTrpSerGluAlaValPheIleHisGluGlyTyrThr 524
Db 1629 GCGGCTCGGAAGAGAACT-----GTGAGAGAAATCATCTCTGCACAAAGCCTT--- 1676
Qy 525 HisAspAlaGlyPheAspAsnAspIleAlaIleuIleLeuAsnAsnAsnValValIle 544
Db 1677 GACCCCAAGACTTACATATGATGATGCTGCTGCGTGAAGATGAAGATGAAGTGTGATG 1736
Qy 545 AsnSerSerHisLeuProIleCysLeuPro-----ArgIleGluAlaGluSerPheMet 562
Db 1737 AATGCTCTTGTGATGCTTGTGCTTCCAGCTCTTCCACAGAGATGAAGAA--CCA 1793
Qy 563 ArgThrAspAspIleGlyThrAlaSerGlyTyrPglYeuThr----- 576
Db 1794 CAGCAAAATCTCTGGATTTGGTAGCGTGCGGCGATTACCAACCCGACACCTGACCTG 1853
Qy 577 -----GlnArgGlyPheLeuAlaArgAsnLeuMetTyrValAspIle 590
Db 854 GAGCATGCCAGTGGCAGTGCAGCAGGCAACACTCTCAAAACATCTTACAGTATGTGAACCTT 1913
Qy 591 ProIleAlaAspHisGlnGlySerCysThrAlaAlaTyrGlu-----LysProProTyrPro 608
Db 1914 CCGGTACACCTGCAGGCTGAGTGCAGAGAGCAGTACAGTACCTTCTGTGATGACTAC-- 1970
Qy 609 ArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGlyLysAspSer 628
Db 1971 -----AATGTGACAGATATACATGTTCTGCGCTGCGCTTCTATGAAGTGGGAAAGACACA 2024
Qy 629 CysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrArgTyrPhe 648
Db 2025 TGCTAGGAGGACAGTGGAGGCGCATTCATATGATGATGATGATGATGATGATGATGATG 2084
Qy 649 ValGlyGlyLeuAlaSerTyrPglYeuSer---MetAsnCysGlyGlyAlaGlyGlnTyrGly 667
Db 2085 GCTCAAGGCGTGTGCTGGTGGTGGTCTGCTGAGAAGGTGGCGATAAACCGGATATATGGA 2144
Qy 668 ValTyrThrIleValIleAsnTyrIleProTyrIleGluSerIleIleSerAsp 685
Db 2145 GTTTACACCAAGATTTCCAAATATGCAAGATGCGCTGCGCAATATAAATGACACAT 2198
RESULT 24
Db 83276 2550 bp mRNA linear VRT 16-NOV-1999
LOCUS D83276
DEFINITION Xenopus laevis mRNA for mannose-binding protein-associated serine
proteinase (MASP), complete cds.
FEATURES
ORIGIN
1 D83276.1 GI:6429054
mannose-binding protein-associated serine protease (MASP).
Xenopus laevis liver cDNA to mRNA, clone_jlb:Xenopus liver cDNA
clone:XA3.7-4-1, race4.
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (sites)
Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H.,
Matsushita, M., Nonaka, M. and Fujita, T.
Two lineages of mannose-binding lectin-associated serine protease
(MASP) in vertebrates
J. Immunol. 161 (9), 4924-4930 (1998)
JOURNAL
MEDLINE 2 (bases 1 to 2550)
99008558
AUTHORS Endo, Y.
TITLE Direct Submission
Submitted (10-JAN-1996) Yuichi Endo, Fukushima Medical College,
Dept. of Biochemistry, 1-Hikarigaoka, Fukushima-shi, Fukushima
960-12, Japan (E-mail: yendoecc.fmu.ac.jp, Tel: 81-245-48-2111,
Fax: 81-245-48-6760)
FEATURES
1. Location/Qualifiers
1.2550
/organism="Xenopus laevis"

/db_xref="taxon:8355"
/clone="XA3.7-4-1, race4"
/feature="type="liver"
/clone_jlb="Xenopus liver cDNA"
1.2550
/gene="MASP"
/gene="MASP"
128..2224
/gene="MASP"
/function="protease"
/codon_start=1
/evidence="experimental"
/protein="BAA8689.1"
/protein_id="BAA8689.1"
/db_xref="GI:6429055"
/translation="MRIPLEFSLCMLLSEAEVIOITDMEGRISLFPDPSYSPSEV
TWNIVPGRGSLKLYFMHDLPSYLCEYDAKVSPOVLANPGRKSTDEQAR
QIITSPSNLSTLSPSDSNERLFPDPAHSALIDIDECTEKSDSDVLDHHCNHTIG
GTYCSRFGLYLDHNTNRCVCSNLFORSGLISSPDYSPVAKSSDCRYLEIE
GFVINDHEDNDVEDHPEVKCPDYLIKTKGNFGLCGERSGRKETSNTVOIL
FHRVYSGENGWRSLYSVTGMPCLHPMPNKLPPQSEYTFKQVYISNOGYRL
KDNVEMESLOIECRKDGTMSSNOIPVOIYDCKKPKREIENGFTTSTAERNRTFOSSFN
YSCREPPYMMVNTILYVTCDSAGEMTQOEGAKIPQCPGCVPRFSRSLARIAGG
KTAKEGISPWIMESDSQNPFCGALISNKWITVAHCHHEIDETDILNSLKW
ELSSRKVILGKHRTLKDDTEQTFQAKNLLHPNPKPTFRFDIALVELSDAFLNDY
VMPICLPEKQVODEHVIYSGWKHFLRLPDSLSLMEVEIPVVGOTLCRTVYOTLELDY
TDEMICAGKEGKDACSDSGSPVTKNELKHHYLAGTVSMVGCGKKIRYGYSD
YKKNLDWIKKXGVOY"
128..178
/gene="MASP"
BASE COUNT 790 a 507 c 581 g 672 t
ORIGIN
sig_peptide
Alignment Scores:
pred. No.: 2.4e-109 Length: 2550
Score: 1492.00 Matches: 289
Percent Similarity: 59.20% Conservative: 113
Best Local Similarity: 42.56% Mismatches: 241
Query Match: 39.70% Indels: 36
DB: 5 Gaps: 11
US-09-874-198-2 (1-686) x D83276 (1-2550)
Qy 26 ValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnArg 45
Db 200 ATGTTGGGGAATCCGCGGTCTCTCTTCCGACTGTATGTCGACGACGACGAGG 259
Qy 46 ArgTyrThrLeuAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAsp 65
Db 260 ACATGGAAATATCACTGTGCCAAGGGGTTTCCGTAAGCTCTATTTTCATGCACTTGTAT 319
Qy 66 LeuGluLeuSerHisLeuGlyGluTyrAspPheValIleLeuSerSerGlyAlaVal 85
Db 320 CTGGAACCGCTTACTCTGCGGATGATGATGATGATGATGATGATGATGATGATGATG 379
Qy 86 LeuAlaThrLeuGlyGlnGluSerThrAspThrGluArgAlaProGlyLysAspThr 105
Db 380 ATTGCAACACTTTCGCGGAAGAAGAACACTGACACAGAACAGCTCAGTCGACAGAT 439
Qy 106 PheTyrSerLeuGlySerLeuAspIleThrPheArgSerAspTyrSerAsnGlyLys 125
Db 440 ATTACATCACTAGTAAATTTTCAACCCCTCACCTTCGCTGATTTCCCAATGAGGAG 499
Qy 126 ProPheThrGlyPheGluAlaPheTyrAlaAlaGluAspIleAspGluGlnValAla 145
Db 500 CGCTTACAGGGTTTATGCTCATTTATGCGCATGATGATGATGATGATGATGATGATG 559
Qy 146 ProGlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuAlaGlyPheTyrCys 165
Db 560 AGTGATGAAGATCTGTTGTGACATCTGACACAAATTAATTAATTAATTAATTAATTA 619
Qy 166 SerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuGlySer 185

Db 620 TCCGACGATTGGTGTACTCTTGCACACGAGTAAACAGAACATGCAAGTGAATGCAGT 679
 Oy 186 GYGLVAlaIhePhrGlnArgSerGlyGluLeuSerSerProGluIyrrProArgProTyr 205
 Db 680 GATPATCTTTTACCACAAAGAGGGGTAAATTCAGATCCGACATCCAGTCCATAT 739
 Oy 206 ProLysLeuSerSerCysThrIyrrSerIleSerLeuGluGluIyrrPheSerValIleLeu 225
 Db 740 GCCAAGAGCTCAGACGTGCATATAGATTTAGAGCTAGACAGAGTTTGTATTAATCTC 799
 Oy 226 AspPheValGluSerPheAspValGluThrHisProGluThrLeuCysProTyrAspPhe 245
 Db 800 CATTTATGATCAACATTTGATTTGTAAGATCACCAGCAATGCAACATCCCTTATGATTAC 859
 Oy 246 LeuIyrrIleGlnThrAspArgGluGluHisGlyProPheCysGlyIyrrThrLeuProHis 265
 Db 860 TTTAAAGTTAAAACTGCAAAAACGAATTTGGCCACTGTGTGAGAGAAATCCGCCGA 919
 Oy 266 ArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGluSerGlyAsp 285
 Db 920 AGAAAGAAACAGAGAACACACGCTTCCATCTCTCCATAGTCAATTCGGGGAA 979
 Oy 286 HisThrIyrrIyrrLysIleHisIyrrThrSerThrAlaGlnProCysProTyrPrometAla 305
 Db 980 AATGAGAGATGAGATGCTCTCTCTGTAACAGGTATGCCATGTCCAAAATCTGCACCCG 1039
 Oy 306 ProProAsnGlyHisValSerProValGlnAlaLysIyrrIleLeuLysAspSerPheSer 325
 Db 1040 CCAATGATATGAAACATTTGAACCTCTCAGCTGATATATACATTTAAAGATCAGGTTGTA 1099
 Oy 326 IlePheCysGluThrGlyIyrrGluLeuGlnGlnIyrrHisLeuProLeuLysSerPheThr 345
 Db 1100 ATTAGCTGCAACCAAGGGTACCGAGTCTGAGAGATTAATGTGAGATGAGAGGTTTACAG 1159
 Oy 346 AlaValCysGlnLysAspGlySerIyrrAspArgPrometProAlaCysSerIleValAsp 365
 Db 1160 ATTGAAATGTGAAAAGATGGAACATGAGTAACCAAAATTCACCTGTTCATTAATGTTGAC 1219
 Oy 366 CysGlyProProAspAspLeuProSerGlyArgValGluIyrrIleThrGlyProGlyVal 385
 Db 1220 TGCAAAAGCCAAAGAGATGAGATGATTCATCTCACTTACCTACAGCAAGAAACCGA 1279
 Oy 386 ThrThrIyrrLysAlaValIleGlnIyrrSerCysGluGluIyrrPheIyrrMetLysVal 405
 Db 1280 ACCCATTCACGTCACGCTTAATCTTCTGCGAGGACACTTACTCATGATGATGTCCTCA 1339
 Oy 406 Asn---AspGlyIyrrIyrrValCysGluAlaAspGlyPheIyrrPheSerLysGlyGlu 424
 Db 1340 AACATAACACTGTGTATCTTGTGATGCAAGCGAGAGATGCACTACCGACAGAGATTGCA 1399
 Oy 425 LysSerLeuProValCysGluProValCysGlyLeuSer-----AlaArgThrThrGly 442
 Db 1400 GCAAAAGTTCTCTACTTCTGACGCCAGTGTGGGGTGCACGCTTTTCCAGAGTGTCTCTA 1459
 Oy 443 GlyArgIleIyrrGlyGlnLysAlaLysProGlyLysPheProGlnValLeuIle 462
 Db 1460 GCAAGATTCAGGTGGAGAGACAGCAAGAGAGATCCCTCGATTTGCAATGTTT 1519
 Oy 463 LeuGlyIyrrIyrrIyrr-----AlaAlaGlyAlaLeuLeuIyrrAspAsnIyrrVal 478
 Db 1520 AGTGACTCCCAAAACATCAAGCCCTTCTGTGAGAGTGCACCTGATTAACCAAAATGATTT 1579
 Oy 479 LeuThrAlaIleHisAlaValIyrrGluGlnLys----- 489
 Db 1580 GTTACAGCTGCCCTCTTGTGACACGAACTGGATACGGAAGACACTGATTAAACTCT 1639
 Oy 490 -----HisAspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSer 506
 Db 1640 TTGAAATGTTTGAACCTCATCTTTCAAAAGTATATTAGTAAACACAGACTTTTAAA 1699
 Oy 507 ProHisIyrrThrGlnAlaIyrrPheSerGluAla-----ValPheIleHisGlyGlyIyrrThr 524
 Db 1700 AAGATGACACAGAGCAAACTTTCAGAGTAAGAACTCATTTCCATCTTAACACTCAAG 1759

Oy 525 HisAspAlaGlyPheAspAsnIleAlaLeuIleLysLeuAsnLysValIle 544
 Db 1760 CCCAANAACA---TTCGCTTTTGACATTTGCCCTAGTGAGCTTTTCAGACAACTTTTCTC 1816
 Oy 545 AsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThr 564
 Db 1817 AATGATTATGTGATGCCAATTTGCTGCTGCGGAGAAACAACTCCACAA----- 1864
 Oy 565 AspAspIleGlyThrAlaSerIyrrGlyLeuThrGlnArgGlyPheLeuAlaArg--- 583
 Db 1865 GATGAGCATGTGATTGTCACGCGATGGGC-----AAGCACTTCTCGAAACGATTA 1915
 Oy 584 -----AsnLeuMetIyrrValAspIleProIleValAlaPheHisGlnLysCysThrAlaAla 601
 Db 1916 CCAGATTTCTCATGAGAGGTGGAATTCGTGTGGGCCACAGACTTGTCCAAAGACTGTT 1975
 Oy 602 TyrcIuLysProProTyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeu 621
 Db 1976 TATCAGACTTTGGAACG-----TTAGTGACTGATGAATGATATATGCGCGGATTC 2026
 Oy 622 GluSerGlyLysAspSerCysArgGlyAspSerGlyAlaLeuValPheLeuAsp 641
 Db 2027 AAGCAAGAGAGAAAGATGCTGCTGTCTGATTTCTGAGAGGCCAAATGCTCAAGAAC 2086
 Oy 642 SerGluThrGluArgThrPheValGlyIyrrIleValSerIyrrPheLysSerMetAsnGly 661
 Db 2087 GAGCTGAAGAAACCTGTGATCTTGCCGGAACGTGTGCATGGGAA---GTGGCTGTGGC 2143
 Oy 662 GluAlaGlyGlnIyrrGlyValIyrrThrLysValIleAsnIyrrIleProTyrIleGlu 680
 Db 2144 AAGAAGATTAAGTATGAGATGATATTCAGATGTTTACAAAGAACCTGCAGCTGATCAAA 2200
 RESULT 25
 AB009073
 LOCUS
 DEFINITION
 AB009073 cyrinus carpio mRNA for mannose-binding protein-linear VRT 14-NOV-1999
 ACCESSION
 AB009073
 VERSION
 AB009073.1 GI:6407537
 KEYWORDS
 MASP; mannose-binding protein-associated serine protease.
 SOURCE
 Cyrinus carpio adult liver tissue_l1b:ZAPPII cDNA cDNA to mRNA,
 clone:2A6.
 ORGANISM
 Cyrinus carpio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Cyrinus.
 REFERENCE
 1 (sites)
 Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H.,
 Matsushita, M., Nonaka, M. and Fujita, T.
 Two lineages of mannose-binding lectin-associated serine protease
 (MASP) in vertebrates
 J. Immunol. 161 (9), 4924-4930 (1998)
 TITLE
 JOURNAL
 MEDLINE
 9908558
 2 (bases 1 to 3229)
 REFERENCE
 Endo, Y.
 Direct Submission
 Submitted (21-NOV-1997) Yuichi Endo, Fukushima Medical College,
 Dept. of Biochemistry, 1-Hikarigaoka, Fukushima, Fukushima 960-12,
 Japan (E-mail:yendo@cc.fmu.ac.jp, Tel.0245-48-2111(ex.2233),
 Fax:0245-48-6760)
 FEATURES
 source
 location/Qualifiers
 1..3229
 /organism="Cyrinus carpio"
 /db_xref="taxon:7962"
 /clone="2A6"
 /tissue_type="liver"
 /dev_stage="adult"
 /tissue_l1b="ZAPPII cDNA"
 16..2253
 /note="MASP"
 /codon_start=1
 /product="mannose-binding protein-associated serine

	protease"	
	/protein_id="naa8686.1"	
	/db_xref="GI:5407538"	
	/translation="MELIRVILVILAQCWPLMTQVILHLTDIYGIKSPNEPESYKPEI	
	/translacion="MELIRVILVILAQCWPLMTQVILHLTDIYGIKSPNEPESYKPEI	
	D/LQNIIVPDQOIRLYFMHDIPIRYSLCEYDYLAKVYSDSEILVAFGKENTDERVP	
	ADNYTSPRNVLSVAFRSDSENERYSGFENHFGAAVUDERDNDRIODLHHCNHY	
	IGGVYSCRGVFLHSRNDKCKVECNMSMYTERGSETLSADPEPYKTSQCYHIL	
	EGEQTILPDDPTDIEDIHPEVNCPYRPIKIHADKDGPCGSGSKITQTSNIVN	
	ILRHSIDTEGEMLKRLTYTDSHPEVCSPLAFLNHLPLRLOSNIYFKRHITLTCRGS	
	LROCKEPIHIOICQRDGKWSVDPFLKINQDGPVVDVLEVIFESFGNSTYFGSKT	
	QYSCRDSPQNNITYTCHQSGEWSVEDTLPETLCIPGEFTLVSNAASQLPTLPSTP	
	LACEOSQLPVAAQKRIYGVGRTASPGELPMTVLISVDVSIRVDRMFSGGALLSTW	
	VLTAFAHLRSHRDFRSVYPAVASEHIRHMLGITLDRIKHLNRSVAVILHPDPON	
	YNNDIILKLSOEYVLSALIOPCVLPBVGNGHMLPTNLGTVAGMGINTANTSAIS	
	TSGLSDLGTSYELLOYVKLPIVPODCEPASYARSNVNTSMPCAGYEGGQDPC	
	LGDSGGAFTVYDADNSGRVNAOGLVSKSGPEEGSQRYRYVYTRNANTILHLMRDE	
	EVARV"	
polya_site	3229	
	/note="25 a nucleotides"	
COUNT	888 a 712 c 692 g 936 t 1 others	
Assignment Scores:		
Pred. No.:	8.08e-107	Length: 3229
Score:	1462.00	Matches: 289
Percent Similarity:	55.00%	Conservative: 118
Best local Similarity:	39.05%	Mismatches: 231
Query Match:	38.90%	Indels: 102
DB:	5	Gaps: 15
US-09-874-198-2 (1-686) x AB009073 (1-3229)		
OY	20	ProLysTPProGlu-----ProValPheGlyArgLeuAlaSerProGly 34
	::::	:::: ::::
Db	61	CCCTTGAGACTCAATGCACCTCCACAGACATATATGGACCATCAAGCTCCAAAC 120
OY	35	PheProGlyGluTyrAlaAsnAspGlnGluArgGlyPheThrAlaProProGly 54
	::::	::::
Db	121	TTTTCTGATCTATACCAAAAGATTAATCATACAGTGGAAATATACGTACCAATGGA 180
OY	55	TyrArgLeuArgLeuTyrPheThrIshPheAspLeuGluLeuSerHisLeucGlyTyr 74
	::::	:::: ::::
Db	181	TATCAAAATCCGAGCTATTATTATATGCACTTCGACATAGAACCCGCTACTATTATGATAGAC 240
OY	75	AspPheValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGluSer 94
	:::: ::::	::::
Db	241	GACTACTGGAAAGGTATACTACAGACTCTCAAGACGCTGCTGTTTGGGAAAGACAAT 300
OY	95	ThrAspThrGluArgAlaProGlyLysAspThrPheTyrSerLeuGlyGlySerSerLeuAsp 114
	::::	::::
Db	301	ACAGCAGCAGAAAGGGTCTCTGCTACATATGATCATACAGTCCGCCAGGAAAGCTGCTCAGT 360
OY	115	IleThrPheArgSerAspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyr 134
	:::: ::::	:::: ::::
Db	361	GTCGCAATTCCTCCGTCAGACTTTCTTAATGAGGAACCTCACTCGGCTTTGAGGCTCAGCTTT 420
OY	135	AlaAlaGluAspIleAspGluGlnValAlaAlaProGlyGluAlaProThrCysAspHis 154
	::::	::::
Db	421	ACTGGGCGCTGATGATGATGAATGCGAAGACAGAAATGATCATCACAGACAGACCTTCATCAT 480
OY	155	HisCysHisAsnHisLeuGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHis 174
	:::: ::::	:::: ::::
Db	481	TTCTCTCAACAACATACATTGGTGGCTTTCACAGTGTCTTCGCCGTATAGGGTCTCTGCTCAT 540
OY	175	ArgAsnLysArgThrCysSerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGly 194
	::::	:::: ::::
Db	541	TCTGCAACACAGAACTGCAAGATGGAGTGAATGAAACATGATACACGAGACGCTCTGGA 600
OY	195	GluLeuSerSerProGluTyrProArgProTyrProLysLeuSerSerCysThrTyrSer 214
	:::: ::::	:::: ::::
Db	601	GAGATCCACCAAGCGCTGATTTTCTGTAACCCGTAACCCCAAAACCTCTGACTGCACATATACAC 660
OY	215	IleSerLeuGlnGluGlyPheSerValIleLeuAspPheValGluSerPheAspValGlu 234

Db	661	ATTGAGCTGGAGAGGGGTTTTCAATACACCTCTGGAATTTGATGACACCTTTACACTTGA	720
Qy	235	ThrlsProgluThrLeuGysProTyrrAspPheLeuLysIleGlnThrAspArgGlu	254
Db	721	GACCACCCGTGAAGTACACCTGCCCCCTACAGATTTTATCAAAATTCATGACGAGAACAGTT	780
Qy	255	HlsGlyProPheGysGlyLysThrLeuProHlsArgIleGlnThrLysSerAsnThr	274
Db	781	TTTGCCCATTTTGTGTGGTACAGTCAACGAGAAATCAGACAGAGATACATAC	840
Qy	275	ThrIleThrPheValThrAspGluSerGlyAspHisThrGlyTrpLysIleHisThr	294
Db	841	AACATTCCTGTTTACACTGACAGACATGGAGAGAACTTGGGTGGAACCTACCTTAC	900
Qy	295	SerThrAlaGlnProGysProTyrrPrometAlaProProAsnGlyHisValSerProVal	314
Db	901	TCCACCCGTTCTGAATGCTCTCTCTTGGCGACCCCTTAATGACACCTGGAGCCTCG	960
Qy	315	GlnAlaLysTyrrIleLeuLysAspSerPheSerIlePheCysGlyThrGlyTyrrLeu	334
Db	961	CAGTCTACATCACTTCAAAAGATCACTATACCTCACTGACCTGTGACCTGTGATCACTG	1020
Qy	335	LeuGlnGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrp	354
Db	1021	AGACAGGGGGGATTAAGAATTTGACATTTATCAATGATGATGATGACAGGATGGGAAATGG	1080
Qy	355	AspArgPrometProAlaCysSerIleValAspCysGlyProProAspAspLeuProSer	374
Db	1081	AGCAGTGTATGCCGCTGCTGTAATAATGGTATGATGGTCCAGTCTG---GATTAATCTG	1137
Qy	375	GlyArgValGlyTyrrIleThrGlyProGlyValThrThrTyrrLysAlaValIleGlnTyrr	394
Db	1138	GGTGGGATTTTGTGAGACTTTGGAACAGTACTGTGTTGATCCAGATCCAGATCCAGTAC	1197
Qy	395	SerCysGluGlnThrPheTyrrThrMetLysValAsnAspGlyLysTyrrValCysGluAla	414
Db	1198	AGCTGACAGAGAC-----TCTCCCGACGGTCAACAC---ACTTACACCTGTATCAG	1245
Qy	415	AspGlyPheThrPheSerSerLysGlyGlyLysSerLeuProValCysGluPro-----	432
Db	1246	AGTGGGAATGGTGATGTAGAGACGG---ACACCTCTGCCACATTGTCTACAGAGAAAT	1302
Qy	432	-----	432
Db	1303	TTTGAAACAACCTTGAATGTCAATGGCGAATCTACGCTCCACTCCACTAACAGCACT	1362
Qy	433	-----ValGysGlyLeuSerAlaGys-----ThrThrGlyLysArgIleTyrrGly	447
Db	1363	CCTCTGATGTGGTGGAGCATACACAGCTTCCCGACGCCACAGACAGCAAGTAATCTTGGT	1422
Qy	448	GlyGlnLysAlaLysProGlyAspPheProTyrrGlnValLeuIleLeuGlyGlyThrThr	467
Db	1423	GGGAGAACAGCATCTCCCGCTGTCTCCCTGCGACAGTCTCTGCTCACTAGAGATGTG	1482
Qy	468	Ala-----AlaGlyAlaLeuLeuTyrrAspAsnTrpAla	478
Db	1483	TCCCGGTGCCAGAGACCGCTGGTTGGCAGCGGGGCTCTCTCTCTCCCACTTGGGTT	1542
Qy	479	LeuThrAlaAlaHisAlaValTyrrGlnGlnLysHisAspAlaSerAlaLeuAspIleArg	498
Db	1543	CTCACAGCGCGGTACGTTTCAATCATCGCGCGGATTTTTCGTGTGCTCCGTCCCT	1602
Qy	499	MetGlyThrLeuLysArgLeuSerProHisTyrrThrGlnAlaTrpSerGluAlaValPhe	518
Db	1603	-----TCTGAACACATTCGA	1617
Qy	519	IleHisGlnGlyTyrrThr-----HisAspAla-----	527
Db	1618	GTTCCACTTTGGGCTCACACGATATACGGGATTAAGCACTTGCTTACCAACCGGCTGTACG	1677
Qy	528	-----GlyPheAspAsnAspIleAlaLeuIle-----	536

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013893; AAH33893.1; -
DR MEROPS; S01.229; -
DR InterPro: IPR000152; CSX_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00084; sushi; 2.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00180; CUB; 2.
DR PROSITE; PS00186; EGF_2; UNKNOWN_1.
DR PROSITE; PS00187; EGF_Ca; UNKNOWN_1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
DR Hydrolase; Lectin; Protease; Serine protease.
QO SQUENCE; 685 AA; 75516 MW; F5F6A6D52BC7099D CRC64;

Just
now signed

Query Match	81.6%	Score 3065.5	DB 11	Length 685
Best Local Similarity	81.5%	Pred. No. 7.2e-255		
Matches 560	Conservative 49	Mismatches 75	Indels 3	Gaps 3

Qy	1	MRLLTLLGLGCVATPLGPKWMPBPVGRGLASPGFEGEYANDDERMYTLAPPEGRRLY	60
Db	1	MRLLIFGLIMSLVATLLGSKWMPBPVGRGLVSPGFPKRYADHOORSTTLAPPEGRRLY	60
Qy	61	FTTHDELISHLCEYDFVKLLSSGAKVLTATLGOESTDTERAPAGKTFEYSLSSLDITFERSD	120
Db	61	FTTHDELSTAYCEYDFVKLLSGTGVLTATLGGESTDTERAPAGNDTFYSLSPSLKYTFHSD	120
Qy	121	YSNEKPTTGFEAFYAADIDECOVAPGEADPTCDHCHNHNLGGFYCSGRAGVYLHRRKRTC	180
Db	121	YSNEKPTTGFEAFYAADIDECRVSFLGDSVPCHYCHNYLGGYVCSGRAGVYLHQRNKTG	180
Qy	181	SALSCGVYFVQRBSELSPEYPRRYPKLSSCTYSISLEBGFVYLDLVESFDVETPEYL	240
Db	181	SALSCGVYFVGRSGYLSBPPEYPPYKLSCTYSIRLEDEGFSYLDLVESFDVETHEAQ	240
Qy	241	CPYQFLKIQDRREHGPRFCGKTLPRHIREKSNMPTIFVTDGSDHDKGMIHNTSTQPC	300
Db	241	CPYQSLKIQDKHGPRFCGKTLPRRIETDSHKATYTFADDESNNHGMKIHNTSTARPC	300
Qy	301	PYPAPNGGHVSPQAVYILKDSFISPCENGYELLOGLHPLKSFYAVCQKDGSDMRMPA	360
Db	301	PDPAPNGGHVSPQAVYILKDRSVYCKRGFELLOGSVPLKSTYAVCQKDGSDMRMPA	360
Qy	361	CSIVDCGPRDDLPSGRVEYITGPQVTTYKAVIOYSCBEFYTKMKVNDGKYUCEADGFWTS	420
Db	361	CSIIDCGPRDDLPGHVDYITGPEVTTYKAVIOYSCBEFYTFYTMSSN-GKYUCEADGFWTS	419
Qy	421	SKGSKLPCVCEPVCGLSARTTGRIYGGQAKPEDPFWOYVLLIGGTT-AAAGALLYDNWVL	479
Db	420	SKGSKLPCVCEPVCGLSTHTIGGRIVGGQPAKPEDPFWOYVLLIGQTTAAAGALLHDWVL	479
Qy	480	TAAAVAYEOKHDSALDIRGTGLKRLSPHTQAMSEAVFTHEGTYHDAGDNDIALIKLN	539
Db	480	TAAAVAYEKRMASSILIRGILKRLSPHTQAMPEEIFTHEGTYHAGCFDNDIALIKLK	539
Qy	540	NKVYVNSINIPICLPRKEASPMETDIGNASGCGLOREGLANRLMYDIPYVDHQKPT	599
Db	540	NKVYVNSIMVCLPRKEASLMTKTDGTGVVAGGILQOKGLANRLMEVDIPLADHQKPT	599

[illegible]

RESULT 2

ID	Q92338	PRELIMINARY;	PRT;	685	AA.
----	--------	--------------	------	-----	-----

DT 01-MAY-1999 (Tremblé, 10, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN MASP2 OR MASP-2.

OC Eukaryota; Metazoa; C

OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A

RX MEDLINE=99008558; Pubmed=9794427;

RA Nonaka M., Fujita T.;

RT (MSP) in vertebrates.";

CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC TRYPSIN FAMILY.

DR HSSP; P00763; 1DPG

DR MGD; MGI:1330832; Masp2.

DR InterPro; IPR001314; Chymotrypsin

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001254; Ser_protease_Try

DR Pfam; PF00431; CUB; 2.

DR Pfam; PF00089; trypsin; 1.

DR SMART; SM00032; CCP; 2.

SMART; SM00179; EGF_CA; 1.

DR PROSITE; PS00010; ASX_HYDROXYLASE

```
DR PROSITE; PS01186; EGF_2; 1.
```

```
DR PROSITE; PS50240; TRYPSIN_DOM;
DR PROSITE; PS50135; TRYPSIN_DOM;
DR PROSITE; PS50136; TRYPSIN_DOM;
```

Calcium-binding; EGF-like domain

```

SQ SEQUENCE 685 AA; 75490 MW; 4571E1AE0703A70F CRC64;

```

Query Match	Score	DB	Length
81.5%	3062.5	11	685

Matches 558; Conservative 50; Mismatches 76; Indels 3; Gaps

1 MRLTLGLCGSVATPLGPKWPEVFGRLASPGFPGGEYANDQERRWTLTAPPGYRLRLY 60

Db 1 MRLLIFGLLWSLVATLLGSKWPEPVFGRLVSPGFPEKYADHQDRSWTLLTAPPGYRLRLY 60

61 FTHEDLESLCEYDFVKLSGAKVATLCQESTDTERAPGKDTFYSLGSSLDITFRSD 120

Db 61 FTHDELSRYCEYDFVKKSLSGTKVLATLCOGESTDTEQAPGNDTFYSLGSLKVTFSHD 120
QY 121 YSNKEPTEGEAFYAADIDECQVAPGEAPTCDDHCHNHGFGFSCSRAGVLRNKRRTC 180
Db 121 YSNKEPTEGEAFYAADIDECQVAPGEAPTCDDHCHNHGFGFSCSRAGVLRNKRRTC 180
QY 181 SALCSGVFTQSRGESSPEYPRPYRLSSCTYSISLEEGSVLLDVESEFVETHEETL 240
Db 181 SALCSGVFTQSRGESSPEYPRPYRLSSCTYSISLEEGSVLLDVESEFVETHEETL 240
QY 241 CPYDLEKIQDREHGFPCGKTLPHRIETKSNVTYTFVDESQDHTGKIMHTSTQAPC 300
Db 241 CPYDLEKIQDREHGFPCGKTLPHRIETKSNVTYTFVDESQDHTGKIMHTSTQAPC 300
QY 301 PYRPAENGHVSPOAKYIILKDSFICFCEYELLOGLHPLKSTAVCAQKDSMDREMPA 360
Db 301 PYRPAENGHVSPOAKYIILKDSFICFCEYELLOGLHPLKSTAVCAQKDSMDREMPA 360
QY 361 CSIVDCGPPDDLPSGRVEYITGPGVTYKAVIOYSCETEFTYMKVNDGKYVCEADGFMTS 420
Db 361 CSIVDCGPPDDLPSGRVEYITGPGVTYKAVIOYSCETEFTYMKVNDGKYVCEADGFMTS 420
QY 421 SKGSKLPVCEPVCGLSARTTGCRITGQAKPDPFMOVLIIGCTT-AAAGALLIDMWVL 479
Db 421 SKGSKLPVCEPVCGLSARTTGCRITGQAKPDPFMOVLIIGCTT-AAAGALLIDMWVL 479
QY 480 TAAHAYVEOKHDSALDIRMGLTKRLSPHYTOAMSEAVFHEGTHAGFNDIALTKLN 539
Db 480 TAAHAYVEOKHDSALDIRMGLTKRLSPHYTOAMSEAVFHEGTHAGFNDIALTKLN 539
QY 540 NKVINSNIPICLPKREAESFMRTDICTAGSGWLTQRFGLANLNVDPIDYDHOCT 599
Db 540 NKVINSNIPICLPKREAESFMRTDICTAGSGWLTQRFGLANLNVDPIDYDHOCT 599
QY 600 AAYEKPPYKGSVTANMLCAGLESGKDSGDSGALVFLDSETEFVNGGIYSWGSNM 659
Db 600 AAYEKPPYKGSVTANMLCAGLESGKDSGDSGALVFLDSETEFVNGGIYSWGSNM 659
QY 660 CGEAGGYVTKVINYIPMIEINISDF 686
Db 660 CGEAGGYVTKVINYIPMIEINISDF 686
RESULT 3
QYJUS8 PRELIMINARY: PRT: 678 AA.
QYJUS8:
01-OCT-2000 (Tremblrel. 15, Created)
01-OCT-2000 (Tremblrel. 15, Last sequence update)
01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Manose-binding protein associated serine protease-2 precursor
DE (Fragment).
GN MASP-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20469449; PubMed=10913141;
RA Wallis R., Dodd R.B.;
RT "Interaction of manose-binding protein with associated serine
proteases: Effects of naturally occurring mutations.";
RU J. Biol. Chem. 275:30962-30969(2000).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC EMBL: AJ277747; CAB90832.1; -
DR HSP: P00763; IDPO.
DR MEROPS: S01.229; -
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.

Juste from

DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; Sush1; 2.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Protease;
KW Repeat; Serine protease; Signal.
FT NON_TER 1 1
FT SIGNAL <1 12 POTENTIAL.
FT CHAIN 13 678 MANNOSE-BINDING PROTEIN ASSOCIATED SERINE
FT PROTEASE-2.
SQ SEQUENCE 678 AA; 74734 MW; 16680E4A9ADCC05 CRC64;
Query Match 81.38; Score 3054.5; DB 11; Length 678;
Best Local Similarity 80.68; Pred. No. 6,2e-254;
Matches 547; Conservative 54; Mismatches 77; Indels 1; Gaps 1;
QY 8 GLICGSVATPLGPKMPVPGFRLSLSGFPGYADDERMTLTAPGRLRLTYTHFDLE 67
Db 1 GLILSLVATILGSGNMPPEVGRGLVSPFPEKYGHHDRSMTLTAPGRLRLTYTHFDLE 60
QY 68 LSHLCEYDFVKKSLSGAVLTLCOGESTDTERABGKDTFYSLSGSLDITFRSDYSNEKP 127
Db 61 LSYRCEYDFVKKSLSGAVLTLCOGESTDTERABGKDTFYSLSGSLKVTFSHDYSNEKP 120
QY 128 TGFAPFAAEDIDECQVAPGEAPTCDDHCHNHGFGFSCSRAGVLRNKRRTCALCSQ 187
Db 121 TGFAPFAAEDIDECQVAPGEAPTCDDHCHNHGFGFSCSRAGVLRNKRRTCALCSQ 180
QY 181 VFTGRSGFLSSPEYPRPYRLSSCAVINIREEGSTILDEYSESDVEMHPAOCPTSLK 240
Db 248 IQDREHGFPCGKTLPHRIETKSNVTYTFVDESQDHTGKIMHTSTQAPCPYPAP 307
QY 241 IQDREHGFPCGKTLPHRIETKSNVTYTFVDESQDHTGKIMHTSTQAPCPYPAP 300
QY 308 NGHVSPOAKYIILKDSFICFCEYELLOGLHPLKSTAVCAQKDSMDREMPACSIYDCG 367
Db 301 NGHVSPOAKYIILKDSFICFCEYELLOGLHPLKSTAVCAQKDSMDREMPACSIYDCG 360
QY 368 PPDDLPSGRVEYITGPGVTYKAVIOYSCETEFTYMKVNDGKYVCEADGFMTS 427
Db 361 PPDDLPSGRVEYITGPGVTYKAVIOYSCETEFTYMKVNDGKYVCEADGFMTS 419
QY 428 PVCEPVCGLSARTTGCRITGQAKPDPFMOVLIIGCTTAAAGALLIDMWVLTAHAAYE 487
Db 420 PVCKPVCGLSARTTGCRITGQAKPDPFMOVLIIGCTTAAAGALLIDMWVLTAHAAYE 479
QY 488 QKHASALDIRMGLTKRLSPHYTOAMSEAVFHEGTHAGFNDIALTKLNKVTINRN 539
Db 480 KTEAMSSLDIRMGILKRLSPHYTOAMSEAVFHEGTHAGFNDIALTKLNKVTINRN 539
QY 548 ITPICLPKREAESFMRTDICTAGSGWLTQRFGLANLNVDPIDYDHOCTAAYEKPP 607
Db 540 IMPICLPKREAESFMRTDICTAGSGWLTQRFGLANLNVDPIDYDHOCTAAYEKPP 599
QY 608 PRGSVTANMLCAGLESGKDSGDSGALVFLDSETEFVNGGIYSWGSNCEAGCYG 667
Db 600 PRGSVTANMLCAGLESGKDSGDSGALVFLDSETEFVNGGIYSWGSNCEAGCYG 659

QY 668 VYTKVNIPIWNIENISDF 686
Db 660 VYTKVNIPIWNIENINNF 678

RESULT 4

Q90X84 PRELIMINARY; PRT; 643 AA.

ID Q90X84;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE MASP-2 protein (Fragment).

GN MASP-2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;

OC NCBI_TaxID=10116;

OX [1]

SEQUENCE FROM N.A.

STRAIN=FISHER 344;

MDLINE=20054576; PubMed=10586086;

RT "The rat and mouse homologues of MASP-2 and Map19, components of the

manan-binding lectin activation pathway of complement.";

RL J. Immunol. 163:6848-6859(1999).

CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

TRYPsin FAMILY.

CC EMBL: Y18572; CAB65389.1; -.

DR HSSP: P00763; IDPO.

DR MEROPS: S01.229; -.

DR InterPro: IPR000152; Asx_hydroxyl.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR000859; CUB_domain.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF_Ca.

DR InterPro: IPR004825; Ins/IGF/relax.

DR InterPro: IPR001254; Ser_protease_try.

DR InterPro: IPR000436; Sushl_SCR_CCP.

DR Pfam: PF00431; CUB_2.

DR Pfam: PF00084; sushl_1.

DR Pfam: PF00089; trypsin_1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS: PR00277; INSULINB.

SMART: SM00032; CCP_2.

SMART: SM00042; CUB_2.

SMART: SM00179; EGF_CA_1.

SMART: SM00020; TRYP_SPC_1.

DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.

DR PROSITE: PS01180; CUB_2.

DR PROSITE: PS01186; EGF_2; 1.

DR PROSITE: PS01187; EGF_CA_1.

DR PROSITE: PS50240; TRYPsin_DOM; 1.

DR PROSITE: PS00135; TRYPsin_SER; 1.

KM Calcium-binding: EGF-like domain; Glycoprotein; Hydrolase; Repeat;

KW Serine protease.

FT NON-TER 1

SO SEQUENCE 643 AA; 70976 MW; 8164B9C0BF688CBB CRC64;

Query Match 77.2%; Score 2901.5; DB 11; Length 643;

Best Local Similarity 80.7%; Pred. No. 8.4e-241;

Matches 520; Conservative 52; Mismatches 71; Indels 1; Gaps 1;

QY 43 QERRMTLAPPGYRLRYLFTHFDELSHLCEDYFVKLSGAKVLAATCGESTPTERAPG 102

Db 1 QDRSWTLTAPPGFRLRYLFTHFNELSLYRCEDYFVKLTSGKVLATLCGSESTTERAPG 60

QY 103 KDTFYSLGSSLDITFRSDYSEKFTGFEPAYAAEDIDECVAPGAPTCDDHCHNHLGG 162

Db 103 KDTFYSLGSSLDITFRSDYSEKFTGFEPAYAAEDIDECVAPGAPTCDDHCHNHLGG 162

Db 61 NDTFYSLGPSLKYTHSDYSEKFTGFEPAYAAEDIDECRTSLGSDYCHNHLGG 130

QY 163 FYSCRAGYVLRHNRKPTCSALSGQVFTORSSELSPEDRPYPKLSCCTYSTSLSEGF 222

Db 121 YVSCRGYVLRHNRKPTCSALSGQVFTGRSGFLSSPEYFPYPKLSCCAVNRLEGEFS 180

QY 223 VILDFVESFVETHPETLCEYDFELKIOTDREHGFPGCKTLPHRIETKSWTVTTFVTE 282

Db 181 IITDFVESFVEMHPEACPYDSLKIOTDKREYCPFGCKTLPHRIETKSWTVTTFVTE 240

QY 283 SGGHTGKRIHTSTAOCPTPMAPRNCHVSPYQAKYTLKRSFSIFCTGELLQHLPLK 342

Db 241 SGNHTGKRIHTSTAOCPTPMAPRNCHVSPYQAKYTLKRSFSIFCTGELLQSVPLK 300

QY 343 SFTAVCKDGSMDRPMACSIIVDCGPPDDLPSCGVEYITPGVTTYKAVIOYSCSEFTYT 402

Db 301 SFTAVCKDGSMDRPMACSIIVDCGPPDDLPNGHVITITPEVTTYKAVIOYSCSEFTYT 360

QY 403 MKVNDGKYVCEADGFWTSKGEKSLPYCEPVCGLSARTTGGRIYGGOKAPGDFPMQVLI 462

Db 361 MSSN-GKYVCEADGFWTSKGEKSLPYCKPVCGLSTHTSGRIITGGOPAKPGDFPMQVLL 419

QY 463 LGGTTAGALLYDNWVLTAAHAYEOKHDSALDIRNGTILKRLSPHTQWSEAVFTHEG 522

Db 420 LGETTAAAGALIHDMVLTAAHAYVGTKEAMSSDIDIRNGILKRLSLYTQAMPFAVFTHEG 479

QY 523 YTHDAGFDNDIALIKLNKKVINSNITPICLPKREASEFMRTDITGASGGLTORGEFLA 582

Db 480 YTHDAGFDNDIALIKLNKKVINSNITPICLPKREASEFMRTDITGASGGLTORGEFLA 539

QY 583 RNLMFYDIPIVDHQKCTAAEKEPPRGVSTANMLCAGLESCKDSCRGDSGALVFLDS 642

Db 540 RNLMFYDIPIVDHQKCTAAEKEPPRGVSTANMLCAGLRGDSCKDSCRGDSGALVFLDN 599

QY 643 ETERMFYGGIVSGSMNCGAGQYGVYTKVNIPIWNIENISDF 686

Db 600 ETORMFYGGIVSGSINCGSEQYGVYTKVNIPIWNIENINNF 643

RESULT 5
Q9JUP3 PRELIMINARY; PRT; 623 AA.

ID Q9JUP3;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE MASP-2 protein (Fragment).

GN MASP-2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;

OC NCBI_TaxID=10116;

RN [1]

SEQUENCE FROM N.A.

STRAIN=FISHER 344;

MDLINE=20054576; PubMed=10586086;

RT "The rat and mouse homologues of MASP-2 and Map19, components of the

manan-binding lectin activation pathway of complement.";

RL J. Immunol. 163:6848-6859(1999).

CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

TRYPsin FAMILY.

CC EMBL: Y18564; CAB70973.1; -.

DR HSSP: P00763; IDPO.

DR MEROPS: S01.229; -.

DR InterPro: IPR000152; Asx_hydroxyl.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR000859; CUB_domain.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF_Ca.

DR InterPro: IPR001254; Ser_protease_try.

DR InterPro: IPR000436; Sushl_SCR_CCP.

DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00084; sushi; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Calcium-binding: EGF-like domain; Glycoprotein; Hydrolase; Repeat;
KW Serine protease.
FT NON_TER 1
SO SEQUENCE 623 AA; 68568 MW; 233AF39E890624F CRC64;
Query Match 74.6%; Score 2802.5; DB 11; Length 623;
Local Similarity 80.6%; Pred. No. 2.7e-232;
Matches 503; Conservative 50; Mismatches 70; Indels 1; Gaps 1;
QY 63 HFDLHSLCEYDFVKKSSAKVLAATLGGESTDTERAPKDTFFSLGSSLDITFRSDYS 122
D 1 HFNELSYRCEYDFVKKLSGTVLATLGGESFTDERAPGNDFFSLGSLKVTFRHSDYS 60
QY 123 NEKPTGFEAFYAAEDIDECOVAGEARPTGDHCHNHLGFGYCGCRAGYVLRHKKRTCSA 182
D 61 NEKPTGFEAFYAAEDIDECRTSLGSDVPCDHYCHNHLGEGYCCRGVYLHONKHRTCSA 120
QY 183 LCGSQVFTORSGLSPSPRYPKLSCTYSISLEGFYSYLDVFSFVETHPETLCP 242
D 121 LCGSQVFTGSGFLSPSPRYPKLSCTYSISLEGFYSYLDVFSFVETHPETLCP 180
QY 243 YDLFKIOTDREHNGPFCGKTLPHRIETKSNVTITFTVDSGDTGKINHSTAQPCPY 302
D 181 YDSLKTIOTDREHNGPFCGKTLPHRIETKSNVTITFTVDSGDTGKINHSTAQPCPD 240
QY 303 PMAPNGHVSVOAKYLLKOSFSIFCEYGLLOGLPLKSFVAVCOGDSWDRPMACS 362
D 241 PTAPNGHISVOAKYLLKOSFSIFCEYGLLOGLPLKSFVAVCOGDSWDRPIECS 300
QY 363 IYDGPDDLPSCGREYITGPGVTTYKAVIOYSCSEFTYTMKVDKGYVCEADGFWTSSK 422
D 301 IIDGPPDDLPNGHVDITTEPEVTYKAVIOYSCSEFTYTMKVDKGYVCEADGFWTSSK 359
QY 423 GKSLLPVCEPVCGLSARTTGRIYGGQAKRPGDFPWQVLLIGTTAAGALLYDMWVLTAA 482
D 360 GKSLLPVCKEPCVCGLSARTTGRIYGGQAKRPGDFPWQVLLIGTTAAGALLYDMWVLTAA 419
QY 483 HAYEOKHDSALDIRMGTLKRLSPHYTOAMSEAVFIHEGTHAGDNDALIKLNKV 542
D 420 HAYGKKEAMSSLDIRMGTLKRLSLITYTOAMPEAVFIHEGTHAGDNDALIKLNKV 479
QY 543 VINSNTPIPLPKREAESFKRTDDIGTAGSGTLORGLARNLAMYVDIPYDHOKCTAAV 602
D 480 TINNMIPICLPKREASLMTKDTDFVTAAGLQKGLARNLAMFVDFPIVDHOKCTATV 539
QY 603 EKPPYPGASTANMLCGLSGCKDCRGDSGALVFLDSETERWVNGVGIWSSMNCGE 662
D 540 TKPPYPAKTYVMNLCAGLRGGCKDCRGDSGALVFLDSETERWVNGVGIWSSMNCGE 599
QY 663 AGQGYTKVINYIPWENIISDF 686
D 600 SEQGYTKVINYIPWENIISDF 623
RESULT 6
Q9OX90 PRELIMINARY; PRT; 541 AA.
AC Q9OX90;
DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE MASP-2 protein (Fragment).
GN MASP-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER 344;
RX MEDLINE=20054576; PubMed=10586086;
RA Slover C.M., Thiel S., Lynch N.J., Schaeble W.J.;
RT "The rat and mouse homologues of MASP-2 and Map19, components of the
mannan-binding lectin activation pathway of complement.";
RL J. Immunol. 163:6848-6859(1999).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC EMBL: Y18566; CAB65383.1; -.
CC HSSP: P00763; IDPO.
DR MEROPS: S01.229; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00084; sushi; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW EGF-like domain; Glycoprotein; Hydrolase; Serine protease.
FT NON_TER 1
SO SEQUENCE 541 AA; 59323 MW; 51BEC0C1A8549CA8 CRC64;
Query Match 64.6%; Score 2427.5; DB 11; Length 541;
Local Similarity 80.4%; Pred. No. 4e-200;
Matches 434; Conservative 43; Mismatches 62; Indels 1; Gaps 1;
QY 147 GEAPTCDHCHNHLGFGYCGRAGYVLRHKKRTCSALCGOVFTORSGLSPSPRY 206
D 3 GDSVPCHYCHNHLGFGYCGRAGYVLRHKKRTCSALCGOVFTORSGLSPSPRY 62
QY 207 KLSCTYSISLEGFYSYLDVFSFVETHPETLCPYDLFKIOTDREHNGPFCGKTLPHR 266
D 63 KLSCAVYNIRLEEGFYSYLDVFSFVETHPETLCPYDLFKIOTDREHNGPFCGKTLPHR 122
QY 267 IETSNVTITFTVDSGDTGKINHSTAQPCPYTMAPNGHVSVOAKYLLKDSFSI 326
D 123 IETSNVTITFTVDSGDTGKINHSTAQPCPYTMAPNGHVSVOAKYLLKDSFSI 182
QY 327 FCEYGEYGLLOGLPLKSFVAVCOGDSMDRPMACSIVDGPDDLPSCGREYITGPGV 386
D 183 FCKTGFEGLLOGLPLKSFVAVCOGDSMDRPIECSIIDGPPDDLPNGHVDITTEPEVT 242
QY 387 TYKAVIOYSCSEFTYTMKVDKGYVCEADGFWTSSKGEKSLPVCEPVCGLSARTTGRIY 446
D 243 TYKAVIOYSCSEFTYTMKVDKGYVCEADGFWTSSKGEKSLPVCEPVCGLSARTTGRIY 301
QY 447 GGQAKRPGDFPWQVLLIGTTAAGALLYDMWVLTAAVYDQKHDSALDIRMGTLKRLS 506
D 302 GGQAKRPGDFPWQVLLIGTTAAGALLYDMWVLTAAVYDQKHDSALDIRMGTLKRLS 361

	DR	InterPro: IPR000436; Sushi_SCR_CCP.
	DR	Pfam: PF00084; sushi_1.
	DR	Pfam: PF00089; trypsin_1.
	DR	PROSITE: PS50240; TRYPSIN_DOM.1.
	DR	PROSITE: PS00135; TRYPSIN_SER; UNKNOWN.1.
	KW	Hydrolase; Serine protease.
	FT	NON_TER 1
	SO	SEQUENCE 323 AA; 35190 MW; 09A86CA502442817 CRC64;
	Query Match	46.6%; Score 1751; DB 4; Length 323;
	Best Local Similarity	99.7%; Pred. No. 3e-142;
	Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Oy	364	VDCGPPDLLPGRVRYITGPVGVTYTKAVIYQVSCETFFYTMKNDGKYCEADGFMTSSKG 423
Db	1	VDCGPPDLLPGRVRYITGPVGVTYTKAVIYQVSCETFFYTMKNDGKYCEADGFMTSSKG 60
Oy	424	EKSILPEPCPGCLSRRTGGRIYGGOKAKPGDFPMQVILLGGTTAGALLYDNWVLTAAH 483
	61	EKSLPECPGVGLSARTTGGRIYGGOKAKPGDFPMQVILLGGTTAGALLYDNWVLTAAH 120
Db	484	AYEEDKHASALDIMGTLLKRISPHYTOAMSFAVIHESGYTHDAEPNDIALIKLNKKV 543
	121	AYEEDKHASALDIMGTLKRISPHYTOAMSEAVFIHESGYTHDAEPNDIALIKLNKKV 180
Oy	544	INSNTIPICLPKEASEFMRTDDIGTAGSGWLTORGFARLNLMYVDIPVDHQCTAAVE 603
Db	181	INSNTIPICLPKEASEFMRTDDIGTAGSGWLTORGFARLNLMYVDIPVDHQCTAAVE 240
Oy	604	KPPYRGSVTANMLCAGLESGCKSDCRDSGALVFLDSETRRWYVGVIVSGSNMGGA 663
Db	241	KPPYRGSVTANMLCAGLESGCKSDCRDSGALVFLDSETRRWYVGVIVSGSNMGGA 300
Oy	664	GQYGVTYKVINYIPIWIENIIDSF 686
Db	301	GQYGVTYKVINYIPIWIENIIDSF 323
RESULT 9		
O90x85		
ID	O90X85	PRELIMINARY; PRT; 366 AA.
AC	O90X85:	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	MASP-2 protein (Fragment).	
	Rattus norvegicus (Rat);	
	Sukaryaola, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_Taxid=10116;	
RM	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-FISHER 344;	
RA	MEDLINE=20054576; PubMed=10586086;	
RT	Stover C.M., Thiel S., Lynch N.J., Schaeble W.J.;	
RT	"Type rat and mouse homologues of MASP-2 and Map19, components of the	
RT	mannan-binding lectin activation pathway of complement.";	
RL	J. Immunol. 163:6846-6859(1999).	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE	
	TRYPSIN FAMILY.	
EMBL	Y18571; CAB65388.1; -	
DR	HSSP: P00763; IDPO.	
DR	MEROPS: S01.329;-	
DR	InterPro: IPR001314; Chymotrypsin.	
DR	InterPro: IPR001254; Ser_protease_TRY.	
DR	InterPro: IPR000436; Sushi_SCR_CCP.	
DR	Pfam: PF00084; sushi_2.	
DR	Pfam: PF00089; trypsin_1.	
DR	PRINTS: PR00722; CHYMOTRYPSIN.	
DR	SMART: SM00032; CCP_1.	
DR	SMART: SM00020; Tryp_Spc; 1.	
DR	PROSITE: PS50240; TRYPSIN_DOM.1.	

DR	PROSITE PS00135: TRYPSIN_SER. 1.
KM	Hydrolase; Serine protease.
FT	NON_TER 1
SQ	SEQUENCE 366 AA; 39882 MW; F84F5ACFE0EC93468 CRC64;
Query Match	43.5%; Score 1634.5; DB 11; Length 366;
Best Local Similarity	80.9%; Pred. No. 3,8e-132;
Matches 297; Conservative 28; Mismatches 41; Indels 1; Gaps 1:	
OY	320 LKDSFSIFCEHYGELLQGHLPLKSFSTAVCOXDGSDMPACSIYDCGPDDLPBSGRVEY 379
Db	1 LKDSFSYCKCGFKFELLOGSVPLKSFSTAVCOKDKDGSMDRPIPECSIIDCCGPPDDLPGHVDY 60
OY	380 ITTGCVATYYKAVIOYSCSEETFYTMKVNGCKYVCENDGFWTSSKGESLIPCEPVCGLSAR 439
Db	61 ITTGEVTTYKAVIOYSCSETFYTMSSN-GKTYCENDEGWTSKSGESLIPVCKPVCGSLTH 119
OY	440 TTGGRTYVGOKAKPCDPFPWOYLILGGTTAAGALLYDNWVLTAHAHVBOKHDAALDIRM 499
Db	120 TSGGRTIIGGQAPKPDPFPMOYLLGETTAAGALLIHDDWVLTAAHAHVYTEAMSLLDIRM 179
OY	500 GTIKRLSHRYQAANSSEAFIEHGTHHDGPNDIALITLNKKVYINSITPICLPKRFAE 559
Db	180 GILKRLSLITYQAAMEAEAFIEHGTHHGAGFNDDIALILKKNKVTNNRMIMPRKADAA 239
OY	560 SFMRPTDIDGTASGWLTORGFLARNLMYVDIPIYDHOKCTAAAYERPPRPGSVTANMLCA 619
Db	240 SLMTDTFVGYTAGWGTLTKGFLARNLMYVDIPIYDHOKCATAYTKQPIPKAVTYNMMLCA 299
OY	620 GLSEGKDCSKDGSGLVFLDSETERMFVGVISWMSGNCGEAGOVYVKVNIYTFWI 679
Db	300 GLDRGCKSKCDGSDGALVFLDNETORWFVGVISWSGINSINGSDPOGVYTKVTNYIFWI 359
OY	680 ENITSDF 686
Db	360 ENIIINF 366
RESULT 10	
ID	09QXD2 PRELIMINARY; PRT: 367 AA.
AC	09QXD2
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Mannose binding lectin-associated serine protease-2 precursor (Fragment).
GN	MASP2 OR MASP-2.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_Taxid:10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=B6CBAL1/3;
RX	MEDLINE=20054576; PubMed=10586086;
RA	Stover C.M., Thiel S., Lynch N.J., Schwaebler W.J.;
RT	"The rat and mouse homologues of Masp-2 and Mp19, components of the
RT	mammalian-binding lectin activation pathway of complement.";
RL	J. Immunol. 163:6848-6859(1999).
CC	-1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARTY).
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR	EMBL; Y19163; CAB65250.1; -.
DR	HSSP; P00736; IAPQ.
DR	MEROPS; S01.229; -.
DR	MGD; MG1:1330832; Masp2.
DR	InterPro: IPRO00152; Asx_hydroxy1.
DR	InterPro: IPRO00859; CUB_domain.
DR	InterPro: IPRO00561; EGF-like.
DR	InterPro: IPRO01881; EGF-Ca.
DR	InterPro: IPRO004825; Ins/IGF/relax.
DR	InterPro: IPRO00436; Sushi_SCR_CCP
DR	InterPro:

DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00084; sushi; 1.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00179; EGF_CA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;
KW Repeat; Signal.
FT SIGNAL; 1 20 POTENTIAL.
FT NON_TER; 367 367
SQ SEQUENCE 367 AA; 40950 MW; F66FCB29CEAB2B7D CRC64;

Query Match 43.3%; Score 1626; DB 11; Length 367;
Best Local Similarity 80.9%; Pred. No. 2,1e-131;
Matches 292; Conservative 29; Mismatches 40; Indels 0; Gaps 0;

2 RLTLTLGLGCSVATPLPKWPEVPGRLASPGFPGEVANDQERRWTLTAPPGYRLTYF 61
7 QLLTFLGLSLVATLGLSKWPEVFGRLVSPGPEKPADHQDSWTLTAPPGYRLTYF 66
62 THFDLELSHLCEDYDFVRLSSGAKVLTALCGQESTDTERAPGKDFYSLGSSLDITFRSDY 121
67 THFDLELSYREYDFVRLSSGAKVLTALCGQESTDTERAPGKDFYSLGSSLDITFRSDY 126
QY 122 SNEPFTGFEEAFYAEDIDECQVAPGEAPTCDDHCHNHLGFGYCSRAGYVLRHNRKTC 181
127 SNEPFTGFEEAFYAEDIDECQVAPGEAPTCDDHCHNHLGFGYCSRAGYVLRHNRKTC 186
QY 182 ALGSGQVTFSGELSGPEYPRPYPKLSCTYSISLEGEFVILDFVESPDVEHHPETIC 241
187 ALGSGQVTFSGELSGPEYPRPYPKLSCTYSISLEGEFVILDFVESPDVEHHPETIC 246
DB 242 PYDFLKIQTDRREHSPFCGKTLPHRIETKSNVTITFVYDESQHTGKIHVYSTAOP 301
247 PYDFLKIQTDRREHSPFCGKTLPHRIETKSNVTITFVYDESQHTGKIHVYSTAOP 306
QY 302 YPMAPPNHVSVPQAKYILKDSFISFCETGELLQHLPLKSFYAVCOKDSMDRPM 361
307 YPMAPPNHVSVPQAKYILKDSFISFCETGELLQHLPLKSFYAVCOKDSMDRPM 366
DB 362 S 362
QY 367 S 367

11 PRELIMINARY; PRT; 728 AA.

AC 096RS4
DB 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Complement factor MASP-3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21378425; PubMed=11485744;
RA Dahl M.R., Thiel S., Matsushita M., Fujita T., Willis A.C.,
RA Christensen T., Vorup-Jensen T., Densenius J.C.,
RA Masp-3 and its association with distinct complexes of the mannan-
binding lectin complement activation pathway.";
RL Immunity 15:127-135(2001).
DR EMBL: AF284421; AAK84071.1;
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.

DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01187; EGF_CA; UNKNOWN_1.
DR PROSITE; PS01188; TRYPsin_DOM; 1.
DR PROSITE; PS00135; TRYPsin_SER; UNKNOWN_1.
KW Hydroxylase; Serine protease.
SQ SEQUENCE 728 AA; 81860 MW; 09B5297A6C14283A CRC64;

Query Match 42.8%; Score 1609; DB 4; Length 728;
Best Local Similarity 43.0%; Pred. No. 1,6e-129;
Matches 307; Conservative 124; Mismatches 245; Indels 38; Gaps 11;

1 MRLTLGLGCSVATPLGKWP-EPYFGRLASPGFPGEVANDQERRWTLTAPPGYRLTYF 59
1 MRLTLGLGCSVATPLGKWP-EPYFGRLASPGFPGEVANDQERRWTLTAPPGYRLTYF 60
60 YTHFDLELSHLCEDYDFVRLSSGAKVLTALCGQESTDTERAPGKDFYSLGSSLDITFRS 119
61 YTHFDLELSHLCEDYDFVRLSSGAKVLTALCGQESTDTERAPGKDFYSLGSSLDITFRS 120
QY 120 DYSNEPFTGFEEAFYAEDIDECQVAPGEAPTCDDHCHNHLGFGYCSRAGYVLRHNRK 179
121 DYSNEPFTGFEEAFYAEDIDECQVAPGEAPTCDDHCHNHLGFGYCSRAGYVLRHNRK 180
QY 180 CSALGQVTFSGELSGPEYPRPYPKLSCTYSISLEGEFVILDFVESPDVEHHPETIC 239
181 CSALGQVTFSGELSGPEYPRPYPKLSCTYSISLEGEFVILDFVESPDVEHHPETIC 240
DB 240 LCPYDFLKIQTDRREHSPFCGKTLPHRIETKSNVTITFVYDESQHTGKIHVYSTAOP 299
241 LCPYDFLKIQTDRREHSPFCGKTLPHRIETKSNVTITFVYDESQHTGKIHVYSTAOP 300
QY 300 CPYMAPPNHVSVPQAKYILKDSFISFCETGELLQHLPLKSFYAVCOKDSMDRPM 359
301 CPYMAPPNHVSVPQAKYILKDSFISFCETGELLQHLPLKSFYAVCOKDSMDRPM 360
QY 360 ACSIVDCGPPDDLPSSGVEYITGCVTAYKAVIQSCSEETFTYM-KVNDGKYCEADGFW 418
361 ACSIVDCGPPDDLPSSGVEYITGCVTAYKAVIQSCSEETFTYM-KVNDGKYCEADGFW 420
DB 361 TCKIVDCRAGFELHGLITSTRNNLTYSKSEIKYSQDEPYKMLNNNTGITYCSAOGW 420
QY 419 TSKGKESLPVCEPVCGLSARTTG--RIYGGOKAKPGDPQVLLIGGTTA----- 468
421 TSKGKESLPVCEPVCGLSARTTG--RIYGGOKAKPGDPQVLLIGGTTA----- 468
DB 421 MNKVLGRSLPTCLPECGQPSRLSLVKRIIGGNAPGLFPMQALILVEDTSVPNDKW 480
QY 469 --AGALLYDMWVLTAAAVYEQKHDSALDI-RMGITLKRISPHYTQAMSA-----VF 518
481 FGSALLSASWILTAALVLSORSDTVIPVSKENHTVYGLHDVRRKSAVNSAARV 540
QY 519 IHGTYTDAGENDIDAIKLNKRVINSNTTPICLPEKESFARTDITAGSMGLTOR 578
541 LHPDF-NIQYNNHDIALVOEPPVLPHPVCLPLPLEEG-PAPMLLVAGMGISNP 598
QY 579 GF-----LARNLMYVDPIVDHOKTAAVEKPPYPRGSYVANNLCAGLSEGGK 626
599 NRVYDELISSGTRLSVGLVYKLPVPHAECKTSYSRS-GNVSVTENNFCAGYEGGK 657
QY 627 DSCRGDSGALVRLDSETERWFGVYSGWS-MNCGAGGYGYTYKYNITPMI 679
DB 658 DTCLGDSGAFVIFDDLQSRWVVOGLVSMGPEPCGSKQYGYVYTKVSNYDVW 711

RESULT 12
ID 092050
AC 092050; PRELIMINARY; PRT; 733 AA.

FT	CHAIN	22	701	MANNOSE-BINDING PROTEIN ASSOCIATED SERINE
FT				PROTEIN-1.
SEQUENCE	701 AA:	79663 MM:	B5CPD619BD3C63DD CRC64:	
Query Match		41.4%;	Score 1555;	DB 11; Length 701;
Best Local Similarity		42.2%;	Pred. No. 6.7e-125;	
Matches 301;	Conservative 132;	Mismatches 227;	Indels 54;	Gaps
OY	2	RLLTLLGLTCSVATPLGKMPPE--	YVFGRLASPGFGEYANDQERKTLTPAPGRLRL	59
DB	4	RRLLYLHVLCCTL--TEVSAHTVLEIMEFQIOISPGY	PDSPDSEYTNATVYPEEGRQL	62
OY	60	YFTHFDELSTLCEYDVKVCSAKVATLATCGSESTFRAKCDPTFSLGSSLDITFRS	119	
DB	63	YFMHFNLLESSLYCYDVKVETEDQVLATPFCGETTDTQTGTGGEVYLSPGFMSVTFTRS	122	
OY	120	DYSNEKPTGFEAFYAAEDIDECOVABGEAPYCDHCHNHGLGFCSCRAGVYLRKRT	179	
DB	123	DFSNEERTGDAHYAMVADVECKEREDELTSCDHCHNYIGGYCCSFYILLTMDRT	182	
OY	180	CSALSGGVFQNSRGSELSPEYRPAPYKLSGCTSYSLSEGRSVLLIDPESPDVTHRET	239	
DB	183	CRVECSGMLFTQKRTITSPDTPYRPKSECSYTTLDLEGMVYLLQEDIPDIEDHEV	242	
OY	240	LCPYDFLKIQTDRREHSPFCGKTLPHRIETKSNVTYITFVTBESGHTGKHHTSTAOP	299	
DB	243	PCPYDIKIKKSGKVMSPFCGKSESPSTIQSHSIQLFRSDNSGNNRMLRSLYAAAGNE	302	
OY	300	CPRYMAPRNGVSVYQAKYLKLSFSIFGCTGELLQGLPLKSTFPAVCQKDGSDMRMP	359	
DB	303	CPKLIQPPYGGKLEPSQAVVSFKQVLLISCTQGIKVLKNDENVMDTQIIECLKDGAWSNKRIP	362	
OY	360	ACSIYDCGPRPDLPSGVAEYITGPGVYTKAVYQYSCSEFTFTMKVFN--DGKYVCEADGEW	418	
DB	363	TCKIIVDCGPAVLKHLGLVSTFRNNLTYYTKSEIRYSCQPYKMLHNTGYVTSAGHGTW	422	
OY	419	TSSGKESLPVCEVYGLS--ARTTGGRIYGGKAKRPGFPMQVLL--GGTTAAGALL	473	
DB	423	TNEVLRKSLPCLPVCGLPFRFSKRHLSIRLFNGRPAKGTTPYIMLSQLNQPFCGSGLL	482	
OY	474	YDNNVLAARAAVY-----EOKHDAALDIRMGTLKRLSPHYTQAMSE-----	515	
DB	483	GSMNVLAARHCHLHRDPRDEEPLIHNHLSLSPDFK----IMGKIMRRSDEDEQHLHYK	538	
OY	516	AVLFHEGTYHDAGFDNDIALIKNNKVININSITPICLPKRAESFMRDTIGT--ASG	572	
DB	539	HIMHLPY--NPSFTENDGLVELSESPRLNDFPMPCLCEHPSTE-----GIWVAVSG	590	
OY	573	WGLDQRGFLAR--NLWYVDPIVDHOKCTAAVERKPRRG--SYANNNICAGLESGGKDS	628	
DB	591	WG-----KQFLRLPRLNMLEITIPLVNHTCOEAYT----PLGKAVIQDMICGEEKGGKDA	643	
OY	629	CRGDSGALVFLDSETERFWVGIVSGMSNGEAGQVYTKVINYIPIWENI	682	
DB	644	CAGDSGGMVTKDAERDQWYLVGVSMGE--DCGKKDRYGVSYIVPNKDWIOGRV	696	
RESULT 14				
Q9QX91				
AC	Q9QX91;	PRELIMINARY;	PRT;	339 AA.
DT	01-MAY-2000 (Tremblrel. 13, Created)			
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)			
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)			
DE	MASP-2 protein (Fragment).			
GN	MASP-2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER 344;			

MEBLINE-2005454576; PubMed-10596086;
 RA Stover C.M., Thiel S., Lynch N.J., Schwaebel W.J.:
 "The rat and mouse homologues of MASP-2 and MAP19, components of the
 RT mannan-binding lectin activation pathway of complement".
 RL J. Immunol. 163:6848-6859(1999).
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 DR EMBL: Y18565; CAB65382.1; -.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.229; -.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYTOTRPSIN.
 DR SMART: SM00352; CCP; 1.
 DR SMART: SM0020; TRYp_Spc; 1.
 DR PROSITE: PS50240; TRYPsin_DOM; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KW Hydrolase: Serine protease.
 FT NON_TER 1
 SQ SEQUENCE 339 AA; 36950 MM; 83EP50150AB37A7 CRC64;
 Query Match 40.3%; Score 1513.5; DB 11; Length 339;
 Best Local Similarity 80.9%; Pred. No. 8.7e-122;
 Matches 275; Conservative 24; Mismatches 40; Indels 1; Gaps 1;


```

Db 595 GVKGHTLMPJNTLGIAGVINGINTANTASTSGTSLDGLTGVSELDQYVKLIYPODCEA 654
QY 601 AYEKPPYPRGSRVAMNLCAGLESGKDCRGDSCGALVFLDSENERFVGVGIVNGS-MN 659
Db 655 SYASRSV-NYNTITSNMCMCAFYEGSGODTCLGDSGAFVTDARSGRVAAOGLVSWGPEE 713
QY 660 CGEAGQYGVYTKVINYIPWI 679
Db 714 CGSORVYGVYTRVANYIHWL 733

RESULT 18
Q9PS25 PRELIMINARY: PRT: 722 AA.
AC Q9PS25:
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
Manose-binding lectin-associated serine protease.
Lampetra japonica (Japanese lamprey) (Entosphenus japonicus).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
Petrromyzontiformes; Petromyzontidae; Lethenteron.
NCBI_Taxid=94989;
ON [1]
RX MEDLINE=99008558; Pubmed=9794427;
RC TISSUE=LIVER.
RA Endo Y., Takahashi M., Nakao M., Salga H., Sekine H., Matsushita M.,
  Nonaka M., Fujita T.;
  "Two lineages of manose-binding lectin-associated serine protease
  (MASP) in vertebrates."
  J. Immunol. 161:4924-4930(1998).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
DR EMBL: AB009075; BAA86868.1; -.
DR HSSP: P00763; IDPO.
DR InterPro: IPR001152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00042; CCP; 2.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00179; EGF_Ca; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_Ca; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KM Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Lectin;
  Protease; Repeat; Serine protease.
SQ SSQUENCE 722 AA; 79591 MW; 81DFB4F78F540B20 CRC64;

Query Match 38.2%; Score 1435; DB 13; Length 722;
Best local Similarity 41.2%; Pred. No. 1.5e-114;
Matches 296; Conservative 112; Mismatches 254; Indels 56; Gaps 19;
QY 4 LTLGLGCGVAPPLGKPE-----PVGRLASPGFGEVANDQERRWTLTAPGYRLR 58
Db 10 LLLLLLTLLTLCARSVVEHHLMSMHGSLRSSGPKPYRPNAAQQTWIRISVPSGYRVR 69
QY 59 LYFHPDLTSLHCEYFVVALSSAKVLAFLCGESTDTERAPEKDFYSLGSSLDITFR 118

```

```

Db 70 LHFTFEDVETSPGCEYDIYKEVYTNELIAGFCGRONCGT--APRAEBMYTPGNVVDITFV 127
QY 119 SDVSNKPEFTGFEAFVAAEDIDEC-QVAPGEAPTCOHCHNHILGFCVSCRAGYVLHRNK 177
Db 128 SDFSNDERYTFGEYFAVIDIDCEEYARVDELACHNCHNNGYVQCSOYGYVLADDN 187
QY 178 RFSALCSGOVFTQSRGELSSPEYPRPYKLSSTCYSLSEEGFSVLLDVESEFDETHP 237
Db 188 RSCVCECSNHVFTERGELSSPDPRPYRALOCKFTLRLEEGFVSLAOEFDFVETHN 247
QY 238 E-TLCPYDFLKIOTDREHGRFCGKTLPHRIETKSNVTVTTEVYDESQDTHGKIHVTS 296
Db 248 EIDTCYPYDSLIRSGCEWEGPLCGTTLPLDPIETRGFEVEVFRSDTSGOKGNMLHYTAK 307
QY 297 AQPCPYMAPRPNHVSVOAKYLLKDSFISFCETGYLLQ---GHLPKSFYAVCOKDS 353
Db 308 SLPCSHVTAHVHSGIEFVQATYSFKDRVHVESCDIGYRLIOOEDGKEVPATLOVECOKDG 367
QY 354 MDRPMFACSIYDCGPPDDLPSGRVEYITGCVTTKAVIQSCSEEPFYTAKVDKGYCE 413
Db 368 MSREPTFCRIYDCGKPAQPKG---PCDSPDVITYKSETDCGCDIAHFTTEMKNGVFTCN 424
QY 414 ADGEWTSKGEKSLPCEPVCGLSARTTG--RIYGGQAKPDGPMQVLI-----LG 464
Db 425 ESGHWITDPRSGEKVPKCEPVCGEPKHVIMELLRLRIVGGRSARSGNFPMQVLTKFMSGRMLG 484
QY 465 GTTAAGALLVDNMYLTFAAHVYEGKHDSALDIRMCTLRLS-----PHYTOMSE 515
Db 485 G---GGALLDRVNLTAHAVAD--YANETTVYLGSMKRVSLKNDPLSGTOOTY---VD 536
QY 516 AVFTHEGYTH-DAGFDNDIALIKL-NNKVVINSNTIPCIPRKE---AESFMETDIDGTA 570
Db 537 KIISHPEYDPLSTGYDNDIALIRAGAVYMTDSVRIICLPYBGGVNPKLSBNDVAFV 596
QY 571 SGWGLIO---RGLARNLKYVDPIYDHQKCTAA---YEKPPYPRGSRVAMNLCAGLES 623
Db 597 SGWGRTAGTGLAMLADTLQYVDLPVPAQACECERANACKWIAELANSTVENNMCAGYSE 656
QY 624 GGRKSCGDSGGALVFLDSENERFVGVGIVNGSMNGCEAGGVYKVIYIPWIEN 681
Db 657 GGRKSCGDSGGGPIVAV--ODNKFETVGVVSWG-MCCAKPGFVGYRVKYLDMWLD 711

RESULT 19
Q9ET60 PRELIMINARY: PRT: 707 AA.
AC Q9ET60:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Serine protease.
GN Clr.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=C57BL/6;
RA Byun S.Y., Hyun J.H., Hwang H.Y., Ryoo Z.Y., Kim T.Y.;
  "Cloning and Sequencing of a cDNA Encoding a Serine Protease
  RT Homologous to Human Complement C1r Precursor from an Allografted Mouse
  RT Skin and Its Expression in Escherichia coli."
  Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
DR EMBL: AF148216; AAG01898.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.193; -.
DR MGD: MGI1355313; Clr.
DR InterPro: IPR001152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.

```


Db 181 LQKDGSCQACSSSELYTERPSGVYSLERDPYRDLRCNSIRVERGLFVHLKLPDPE 240
QY 233 VETHPETLCPRDFKIQTDREHGRFCGKTLPHRIETKSNVTITFTVDESGDHTGKI 292
Db 241 IDHQVHCPRDQLOITANGKNLGEFCGKORPDLTSSNAVLDLFTDESGRMKLH 300
QY 293 YTSAPQRYPMAPRN-GHVSPOAKYILKDSFSIFCETGYELLQHLPLKSFYAVOC 351
Db 301 YTTETIKCPORALDEFTIIDDPPQYOFRDYFIVTCQKQYQLMEGNALLSFTAVOC 360
QY 352 GSNRPAPACISVDCGPRDLPSCREYITGCPYTTKAVIQYSCETFTYMKVNDCK 409
Db 361 GTWHAPRCKIKKCGQPOSLSNDGFRITTKGVTYTEASIQYCHPEYKMLRACSSSE 420
QY 410 -----YCEADGFWTSSKGEKSLPVCEVCG--LSARTGGRYGGGAKGDPFWOLI 462
Db 421 SMRIYCTAOGIKKNEEGEKMRCLPVCGKRVNPTQKRIIRGAPARGNPPWQFT 480
QY 463 LGTTAAGALLYDWMVLTAAHAYEQRKNDAS-----ALDIRMG-----TLKRLSPHYTQ 511
Db 481 TTHGRGAGALLGDRWILTAHTIYRKHNKENDNANPKMLFELGHTNVEQIKKGNHPRV 540
QY 512 AMSAVYIHEHYTHDA--GENDIALIKLNKVVINSNTPICLPKREAESFMTDICT 569
Db 541 ----RVIIHPDYRQDEBNNEEGDIALLENSYTLGPELLPICLP--DNETFYGQGLMGY 594
QY 570 ASGGLTORGLARNLVYDPIVDHQC-----TAAYEKPRPGSVTANLACGLSSG 625
Db 595 VSGGITE-DKLAFDLFRVLPAVDSQCQWLOTKQDTSF-----SONFEGGDPVQ 648
QY 626 KDCRGDSGALVFLDSETERMEVGGIVSGSMNCGAGQYVYTKYINIPWENIISD 685
Db 649 QDACGDSGGVFAVRDRNRDITWATGIVSMG-IGCGEG--YGFYTKLVNYDWMIKKMGD 705

RESULT 21

Q8BVY4 PRELIMINARY; PRT; 694 AA.
ID Q8BVY4
AC Q8BVY4
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 77.4 kDa protein (similar to complement component 1, s subcomponent).
DE Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthelia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Liver.
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC022123; AAI18319.1; -;
DR EMBL: BC018319; AAI18319.1; -;
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushl_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00084; Sushl; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRPSIN.
DR SMART: SM00032; CCP; 2.
Mus musculus (Mouse).

DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01187; EGF_CA; UNKNOWN_1.
DR PROSITE: PS0240; TRYPsin_DOM; 1.
DR PROSITE: PS00135; TRYPsin_SER; UNKNOWN_1.
KW Hypothetical protein
SQ SEQUENCE 694 AA: 77415 MW: 06AA125D5E8227F5 CRC64;

Query Match 32.8%; Score 1232; DB 11; Length 694;
Best Local Similarity 38.7%; Pred. No. 4.1e-97;
Matches 270; Conservative 99; Mismatches 288; Indels 40; Gaps 18;

QY 1 MRLTLGLGCGSATVATLGKRWPEPVGRGLASPGFGEIYANDERRMTLAPRGRLRLY 60
Db 7 MCVLPLSLASFSAEP-----TMHGELISPNYPOAYPNVDVKSMDIEVEGFCIHLY 59
QY 61 FTHFDLISHLCEYDFYKLSSGAKVLAFLCGSESTDRERAPGKDTFSLSSLDITFRSD 120
Db 60 FTHVDIEPESCAVDSVQIISGIEBRLCGQKTSKSPNSPIIEFQFPYKLOVFTSD 119
QY 121 YSNEKPEGEAFYAADIDECQVAPGEAPTCDHCHNHILGFCSCRAGVYLHNRKTC 180
Db 120 FSNERTFTGFAAYTALDINECTDFT-DVP-CSHFCAANNFTGGYGCSCRPYFLHDDMRNC 177
QY 181 SALCSGOVFTORSGELSSPEYPRPYKLSCTYSISLEEGFSVILDFV-ESFDVE-THPE 238
Db 178 GVNCSDGVFTALICEISSPNPNYPENSRCQYIQLOEGFOVYVYMQREDFVEPADSE 237
QY 239 TLCPYDFLKIQTDREHGRFCGKTLPH--RIETSNVTITFTVDESGDHTGKIHTST 296
Db 238 GNCP-DSLTPRASKNQOQPRPCGNGFPGPLTITQSNLTGLVFOYDLGOKKGMRLRHGD 296
QY 297 AQCPRYPMAPRNGHVSPOAKYILKDSFSIFCETGYELLQHLPLKSFYAVOCGSMR 356
Db 297 PISCARKIT-ANSTWEPDKAKYVFKDVKITCVGFEVGEHVSSTSYISFCQSDGQMSN 355
QY 357 PMRCSIVDCGPRDLPSCREYITGCPYTTKAVIQYSCETFTYMKVND-GKYVEAD 415
Db 356 SGLCQPVYCGIPDPINGAYE---EPENSVFQVIVHYTCEERYVMEHEGEGEYRCAN 412
QY 416 GFWTSSKGEKSLPVCEVCGISAR--TTGGRIYGGOKAKGDPFWOLILGTTAAAL 473
Db 413 GRWVNDQIGLELPCTACGVPPEPVHQRITFGQPAKLENPWQY-FNNHPRASGALI 471
QY 474 YDMNVLTAAHAYEQRKNDASALDIRMGTLLKRLSPHYTQAMSEAVFIHEGYTHD-----A 527
Db 472 NEYVWLTAAH-VLEKISDPLMYVGTMSVRRITLLENAQRLYSKRVFIHPSMKKEDDPNRT 530
QY 528 GFNDIALIKLNKVVINSNTPICLPKREAESFMTDIDIGTASGGLTORGLARNLMY 587
Db 531 NFDNDIALVQLKDPVKMKPKVSPICLPGTSEYVNSGDMGLISGMSTEKKRVILRG 590
QY 588 VDIPIVDHOCATAVEKPRVPGS--VTANMLCAGLESKDCRSDSGALVFL--DS 642
Db 591 AKYVYTSLETCQYKEENPVYRPEDYFTDMTAG--EKGVDSCGSDSGAFAFOVPNV 648
QY 643 ETERMEVGGIVSGSMNCGEAGQYVYTKYINIPWI 679
Db 649 TVPKFYVAGLVSMGK-----RCGTGYGVYTKVKNYDWM 681

RESULT 22

Q8R099 PRELIMINARY; PRT; 694 AA.
ID Q8R099
AC Q8R099
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to complement component 1, s subcomponent.
OS Mus musculus (Mouse).


```

Db 378 ---DPEDTVESVTHYCEEPYTYMEDEEGEYICAAANGSVNDQLGVELKCIIPVCGVP 434
QY 438 AR--TTGRIYGGOKAKPGDEPMQVLI---LGCTTAAGALLYDMVLTAAHAYVEOKHDA 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 TEPKVOGRIGGQSTKIQSPMQVYFESPRG----GALIDEVNLTAANVEGNSDPV 490
QY 493 SALDIRKGTILKRLSPHYTQAMSEAVFIHEGYTHD-----AGFNDIALIKLNKVVINS 546
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 491 MYVSTLTLEKLERLR-NAQRLITERVIIHPSWKQEDDLNTFTNFDNDIALVQLKDPVKMGP 549
QY 547 NIPICLPKRAESFMRTDIDGTASGGLTQGRFLANLMLVVDIPIVDHOKCTAAYEKRP 606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 TVADICLPETFSDDINSEVSDGLISGMRTEIRTNVIOLRGAKLPITSLKCGQGVKVP 609
QY 607 YPRGS---VTANMLCAGLESQKDCRQDSGG--ALVFIDSETERMFVGIVSGSMNCG 661
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 610 KARSNDVYFDNMICAG--EKGVDSCEGDSGAFALPVPNPKDKRFYAGLVSGK---- 663
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
    662 EAGQYVYTKVINYIPMI 679
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
    664 KCGYGIYTKVKNYDWI 681

RESULT 24
09DGC1 ID Q9DGC1 PRELIMINARY; PRT; 685 AA.
AC Q9DGC1
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cirs-B protein.
OS Cyprius carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_taxid=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEPATOPANCREAS;
RX MEDLINE=2115142; Pubmed=11220628;
RA Nkao M. Ohsaka K., Kato Y., Fujiki K., Yano T.;
RT "Molecular cloning of complement C1r/C1s/MASP2-like serine proteases
RT from the common carp (Cyprinus carpio).";
RL Immunogenetics 52:255-263(2001).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
    -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
    TRYPSIN FAMILY
EMBL: AB042610; BAB17846.1; -.
HSSP: P00763; IDPO.
DR Interpro: IPR000152; Asx_hydroxyl.
DR Interpro: IPR001314; Chymotrypsin.
DR Interpro: IPR000859; Cub domain.
DR Interpro: IPR000561; EGF-like.
DR Interpro: IPR001881; EGF_Ca.
DR Interpro: IPR001254; Ser_protease_Try.
DR Interpro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00084; Sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00179; EGF_Ca; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF; 2; 1.
DR PROSITE: PS01187; EGF_Ca; 1.
DR PROSITE: PSS0240; TRYPsin_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.

```

```

KW EGF-like domain; Glycoprotein; Hydrolase; Hydroxylation;
KW Serine protease.
SQ SEQUENCE 685 AA; 76838 MW; 169BBFD8A8A745BD CRC64;
Query Match 30.8%; Score 1158; DB 13; Length 685;
Best Local Similarity 36.3%; Pred. No. 9.3e-91;
Matches 247; Conservative 115; Mismatches 284; Indels 34; Gaps 14;
QY 24 EP-VFGRLASPGFGEAEVANDQERWTLTAPRGYRLRYFTFTDELSHLCEYDFKSSG 82
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21 EPAMFGEVSSFOYQPIRANFQEDMLEVPOGYQIQTLTFNHLDIERSPCYTDYSVSD 80
QY 83 AKVLATLOGESTDTERAPGKDTFVSLGSSLDITFRSDYSNEKPFTEAFYAEDIDEC 142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 81 NKVLGKFCQGQSTDRFH-PGDKPLAPRNLQVFLVDVSNHDLIGFTATYQVADDEC 139
QY 143 QVAFGEAPTCRHHCHNLGFCYSCRACYLHRRKRTCSALCSQVTFORSGELSPETP 202
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 140 SSSSEVNAPCSQICLINTGLSYLCAHHGYMLRPQRTCVLECGGCVHSESGTSSPCFP 199
QY 203 RPYPKLSCTYSISLEBGSYILDFVESFDET--HPEITLCPYDFLKIQTDREHGPFG 260
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 200 DVSPLNLDCTIYTSVQCYMITLNFSONFHEQYVNOGQTCLEFHMLOYVPGKPKETCG 259
QY 261 KTLPHRIETKSNVTYITFEVTDSEGDHTGWMKIHVYTAQPCPYMAPRPNGHVSPYQAKYIL 320
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 260 GKSPGVLTNGANFVQLEHNTDRYQSGQSMIHYTORQCPHPIGILGTTTPNFAQVLY 319
QY 321 KDSFSIFCTGYELLQHLPLKSTFAYVCQKDSMDRPPACSIYDCGPPDLPSGRVEYI 380
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 320 RDYIHVRCPMGYKLMGKEKILSFKISQCSNGKWHLLTLPCKIIDCGAPKOLLNDGFQFI 379
QY 381 TGPQVTTYKAVIYQSCSETEY----TMKVNQKIVCEADGFWTSKSKSLPVECPVGL 436
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 380 SGEN-NEILSVIETRCNPEPYTRFKDTAKV--YRCAVERKMEVNNNDIIPCYPECGM 435
QY 437 SARTT-GGRIYGGOKAKPGDEPMQVLI---LGCTTAAGALLYDMVLTAAHAYVEOKHDA 495
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 436 NTEVSPFGGRVFGQARQGIIPQDLKMHTRGASLSIDVWALTAANVVDGELTMTW 495
QY 496 DIRKGTILKRLSPHYTQAMSEAVFIHEGYT-----HDAGFNDIALIKLNKVVINSIT 549
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 496 LGGITTEARNQMPVMEFA--EKIIHPIKQYRVSRLRGHQTNVNDIALIKMSARVPLGRPIR 553
QY 550 PICLPKRAESFMRTDIDGTASGGLTQGRFLANLMLVVDIPIVDHOKCTAAYEKRPYR 609
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 554 PVLCPNKTHEVME-GYIGTVSGFGKREGTLSEILRTGIREYPEREC--VFGKMP--- 607
QY 610 GSVTANMLCAGLESQKDCRQDSGALVF---LDSETERMFVGIVSGSMNCGEAGQ 665
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 608 --VEKNNMFCAGDDVEHVDSCGDSGGPLFFPMLGHSKEQRYEVRGIVSMGPNVCGDVSK 665
QY 666 YGVYTKVINYIPMIENIISD 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 666 -AYYTKVENYLGWIEETMAN 684

RESULT 25
09DGC0 ID Q9DGC0 PRELIMINARY; PRT; 685 AA.
AC Q9DGC0
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE C1r/s-B protein.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_taxid=7962;
RN [1]
RP SEQUENCE FROM N.A.

```

RC TISSUE=HEPATOPANCREAS;
 RA MEDLINE=2115142; Pubmed=11220628;
 RA Nakao M., Ohsaka K., Kato Y., Fujiki K., Yano T.;
 RT "Molecular cloning of complement C1r/C1s/MASP2-like serine proteases
 from the common carp (*Cyprinus carpio*).";
 RL Immunogenetics 52:255-263(2001).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL: AB042611; BAB17847.1; -.
 DR HSSP: P00763; 1DPO.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00008; EGF; 1.
 DR Pfam: PF00084; trypsin; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS02440; TRYPSIN_DOM; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR EGF-like domain; Glycoprotein; Hydrolyase; Hydroxylation;
 KW Serine protease.
 SQ SEQUENCE 685 AA; 76831 MW; E49B1891D7280F36 CRC64;

Query Match 30.8%; Score 1158; DB 13; Length 685;

Best Local Similarity 36.0%; Pred. No. 9,3e-91;

Matches 244; Conservative 116; Mismatches 289; Indels 28; Gaps 13;

QY 24 EP-VFGRLASGFPGEYANDQERMTLTAPPYRLRLYTHFDELSHLCYDFVKLSG 82
 DB 21 EPAMFGEVSSPQYPPYANFQEQWLEVPQYQIQLTENHLIDIEPSPDYDSVYSVD 80
 QY 83 AKVLAITLGGESTTERAPGKDTFYSLGSSLDITFRSDYSNEKPTGFEAFYAEDIDEC 142
 DB 81 NKVLGKRCGQSTDRFH-PGDKPLAPHNRLQVLFTDVSNHEHLIGFTAFYQAVDIDEC 139
 QY 143 QVAPGEAPTCDDHCHNHLGGFYCSRAGYVLRHNRKPTCSALCSGOVFTORGSELSPPEP 202
 DB 140 SSSSEVENAPCSQICLNLTLGSLCHHGMYLRPQRTCVLECGGVSSESGTISSPGFP 199
 QY 203 RPYRLSSCTYISLEGGFVILDFVESFDVET--HPETLCPYDFLKIOTDREHGPFCG 260
 DB 200 DVSPNLDCITYTISVQGYMTLNFSONFHIEQYNOGQTCLEFHLQVSVGKKPKKYCG 259
 QY 261 KTLPHRIETKSNYTYTFVDESQDHTGKTHYSTAOPCPYMAPPNHGAVSPVQAKYIL 320
 DB 260 GKSPGVINTGANFVQLEHYHNDRYGOSQGSINHYTTORVOCPHNGIIGNCTITPNAQYLY 319
 QY 321 KDSFSICFETGYELLQGLHPLKSTAVCOKDGSWDRPMAPACSIYDCGPPDDLPGRAVEYI 380
 DB 320 RDIYHVCMPEYKILMGEKELISFKSICQSGKKNHLLPBECKITIDCGAPKOLLNGDFOFI 379
 QY 381 TGPQVTTYKAVIYQSCETETFTMK-VNDGKYVEADGFWTSSKGEKSLPYVCEPYCGLSAR 439
 DB 380 SGEN-NEYLSVIEYRCNEPYRFKDTAKATYRCAVDRKWTENVNNNDIIPCYPCGMNTE 438
 QY 440 TT-GGRTYGGOKAKPGDFPQVLLGGTTAAGALLYDNWVLTAAHAYYEQKHDAALDIR 498
 DB 439 VSEFGRVFGGQARPGQIPWOLLHKMHTRGASLISDYWALTAHVVDGLENTMTWLG 498

QY 499 MGLKRLSPHYTQAMSEAVFTHEGYT-----HDAGEFDNDIALIKLNKKVYINSNTPIC 552
 DB 499 ITEARNQNPVYMEA--EKIIHPDYQVSLRGHOTNTYNDNDALIKMARVPDLGNIRPVC 556
 QY 553 LPRREASEFMFTDDIGTASGWLTORGFELARNLMYVDIPIYHQKCTAAVEKPPYPGSV 612
 DB 557 LPNKTHE-FVMEGTMTGTVSGFGFKESLTSEILRIGHIREYASBK--VFGKMP-----I 608
 QY 613 TANMLCAGLESQKDSQSGDGGALVE----LDSETERWVPYGGIVSGSNMCGEAGQYV 668
 DB 609 SKNMFCAQDDVEHVDSCQGSQGPLFPMLGHGSKREQREYRGIVSWGPRCGDVSK-AY 667
 QY 669 YTKVINYIPWIENTISD 685
 DB 668 YTKVENTIGWIEETMAN 684

Search completed: January 23, 2003, 22:19:54
 Job time : 96 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 22:14:16 : Search time 20 Seconds
(without alignments)
1009.206 Million cell updates/sec

Title: US-09-874-198-2
Perfect score: 3758
Sequence: 1 MRLTLGLGCGSVATPLGP.....GYTKVINYPIWENISDF 686

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

ed: 262574 seqs, 29422922 residues
number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 75 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCtUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	480	13.0	855	2	US-09-027-337-2
2	480	13.0	855	4	US-09-644-600-2
3	470.5	12.5	1019	1	US-08-296-014A-4
4	470.5	12.5	1019	2	US-08-596-405-4
5	470.5	12.5	1019	2	US-08-877-620-4
6	470.5	12.5	1083	1	US-08-296-014A-2
7	470.5	12.5	1083	2	US-08-596-405-2
8	470.5	12.5	1083	2	US-08-877-620-2
9	446.5	11.9	269	4	US-09-715-994-2
10	445.5	11.5	242	4	US-09-644-600-10
11	430.5	11.5	798	1	US-08-944-483-29
12	408.5	10.9	798	1	US-08-200-900A-2
13	408.5	10.9	798	5	PCt-US94-00616-2
14	405	10.8	448	1	US-08-295-411-3
15	405	10.8	448	1	US-08-955-471-3
16	405	10.8	448	3	PCt-US92-10242-3
17	405	10.8	461	3	US-08-742-877-2
18	405	10.8	461	4	US-09-053-871A-21
19	405	10.8	461	6	5521070-2
20	404	10.8	415	1	US-08-073-531B-1
21	404	10.7	415	1	US-08-766-288-1
22	403	10.7	448	5	PCt-US92-10068-1
23	401	10.7	1013	2	US-08-866-650-5
24	401	10.7	1013	2	US-09-021-287-5
25	401	10.7	1013	3	US-08-991-408-2
26	401	10.7	1013	4	US-09-240-473-5
27	401	10.7	1013	4	US-09-432-473-2

28	400	10.6	415	1	US-08-295-411-2	Sequence 2, Appl1
29	400	10.6	415	5	US-08-955-471-2	Sequence 2, Appl1
30	400	10.6	415	5	PCt-US92-10242-2	Sequence 2, Appl1
31	400	10.6	437	1	US-08-487-037-2	Sequence 2, Appl1
32	400	10.6	591	3	US-08-991-408-4	Sequence 4, Appl1
33	400	10.6	591	4	US-09-432-473-4	Sequence 4, Appl1
34	399	10.6	488	1	US-08-487-037-1	Sequence 2, Appl1
35	398	10.6	416	2	US-09-000-846-2	Sequence 1, Appl1
36	396.5	10.6	788	1	US-08-872-225-1	Sequence 1, Appl1
37	396.5	10.6	986	4	US-08-872-225-4	Sequence 3, Appl1
38	393.5	10.5	1013	2	US-08-866-650-3	Sequence 3, Appl1
39	393.5	10.5	1013	2	US-09-021-287-3	Sequence 3, Appl1
40	393.5	10.5	1013	4	US-09-021-287-3	Sequence 3, Appl1
41	392	10.4	437	1	US-09-240-473-3	Sequence 3, Appl1
42	391.5	10.4	437	4	US-08-487-037-3	Sequence 3, Appl1
43	389.5	10.4	235	4	US-08-944-483-28	Sequence 28, Appl1
44	389.5	10.4	235	4	US-08-944-483-48	Sequence 48, Appl1
45	389.5	10.4	461	6	5270178-2	Patent No. 5270178
46	387	10.3	461	6	5460953-3	Patent No. 5460953
47	384	10.2	622	3	US-08-952-967-8	Sequence 8, Appl1
48	382	10.2	409	4	US-09-065-872-2	Sequence 2, Appl1
49	382	10.2	410	4	US-09-065-872-1	Sequence 2, Appl1
50	382	10.2	410	4	US-09-065-872-1	Sequence 1, Appl1
51	382	10.2	419	1	US-09-667-570A-1	Sequence 1, Appl1
52	382	10.2	419	2	US-08-295-411-1	Sequence 1, Appl1
53	382	10.2	419	4	US-08-955-471-1	Sequence 1, Appl1
54	382	10.2	419	5	US-09-667-570A-3	Sequence 3, Appl1
55	382	10.2	460	2	PCt-US92-10242-1	Sequence 1, Appl1
56	382	10.2	460	2	US-08-756-506-2	Sequence 2, Appl1
57	382	10.2	460	2	US-08-756-506-4	Sequence 2, Appl1
58	382	10.2	461	6	5225537-2	Patent No. 5225537
59	380	10.1	730	4	US-09-117-708-14	Sequence 14, Appl1
60	377	10.0	579	2	US-08-872-757-2	Sequence 2, Appl1
61	377	10.0	579	2	US-08-295-411-4	Sequence 4, Appl1
62	377	10.0	579	5	US-08-555-471-4	Sequence 4, Appl1
63	377	10.0	615	1	PCt-US92-10242-4	Sequence 4, Appl1
64	377	10.0	615	1	US-07-998-972A-3	Sequence 3, Appl1
65	377	10.0	615	1	US-08-463-953-3	Sequence 3, Appl1
66	377	10.0	615	5	US-08-463-953-3	Sequence 3, Appl1
67	376	10.0	251	4	PCt-US92-11357-3	Sequence 3, Appl1
68	375.5	10.0	306	1	US-08-944-483-47	Sequence 47, Appl1
69	375.5	10.0	306	1	US-08-330-978-1	Sequence 1, Appl1
70	375.5	10.0	306	1	US-08-474-042-1	Sequence 1, Appl1
71	375.5	10.0	306	1	US-08-484-558-1	Sequence 1, Appl1
72	372	9.9	461	6	US-08-774-592-1	Sequence 1, Appl1
73	372	9.9	461	6	5270178-17	Patent No. 5270178
74	370.5	9.9	241	1	US-08-330-978-4	Patent No. 5270178
75	370.5	9.9	241	1	US-08-474-042-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-027-337-2
: Sequence 2, Application US/09027337B
: Patent No. 5972616
: GENERAL INFORMATION:
: APPLICANT: O'Brien, Timothy J.
: TITLE OF INVENTION: Tanimoto, Hirotoishi
: TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed In
: TITLE OF INVENTION: Breast and Ovarian Carcinomas
: FILE REFERENCE: D6064
: CURRENT APPLICATION NUMBER: US/09/027,337B
: CURRENT FILING DATE: 1998-02-20
: NUMBER OF SEQ ID NOS: 13
: SEQ ID NO 2
: LENGTH: 855
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
: :
: OTHER INFORMATION: 23 to 2589 of Sequence 1


```

: GENERAL INFORMATION:
: APPLICANT: Ding, Jeak Ling
: APPLICANT: Ho, Bow
: TITLE OF INVENTION: The Cloned Factor C cDNA of the
: TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinus scaber
: TITLE OF INVENTION: rotundicauda and purification of Factor C Proenzyme
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: 8110 Gatehouse Road, Suite 500 East
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22042
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: URGENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/296.014A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy, Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1781-105P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 205-8000
: TELEFAX: (703) 205-8050
: TELEX: 248345
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1019 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-296-014A-4
:
: Query Match 12.5%; Score 470.5; DB 1; Length 1019;
: Best Local Similarity 28.1%; Pred. No. 2,7e-31;
: Matches 151; Conservative 63; Mismatches 187; Indels 137; Gaps 26
:
: QY 265 HRIET-----KSNIV--TTF-VTDSGDHTG-----WKIHTSTA 297
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DP 500 HRLDEKEPFIMELMDRSNVNLNDNLTFVAGSGEGNETNCVYMDIDOLQSVKT--KSCF 557
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DB 558 QPSSTCAAMDLSDRKAKKADCDPGSLENGHAT--LHGQST--DGFYAGSSIRYCEVYLHYLS 614
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: QY 342 KSTFVAVCKDGSWDRPMAPC-SIVDC-GPPDDLPS-GRVEYITGPGVT----- 387
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DB 615 GTEIVTCTTNGTWSAPKPRCIVKITQGNP--VPSYGSVE-IKPPRTNSISRVGSFPLR 671
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: QY 388 -----YKAVIQ--YSCSEFFITMKVND 407
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DB 672 LPRLLPLAARAKPPPKPRSSOPSTVVDLASKYKLPBGHYRVGSRAIYTESRYELLGSO 731
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: QY 408 GKUYCEADGFMTSSKGEKSLPVCSEFVCGLSARTTGGRIIGCGAKRGDPMOVL----- 462
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DB 732 GRR-CDSNGNMGRBAS-----CIPVCGRSDSPRSPFIIWNGNSTELGQPMQAGISRMLA 785
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: QY 463 ---LGGTTAAGALLYDNNVLTAAHAVYEOKHDASALDIMGTLKRLSPHYTQAMS----- 514
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DB 786 DHNMVFLOCGGSLNLEKMTVTAACHV--TYSATAEIIIDPNQFMKYLKYYRDSRDDDY 842
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: QY 515 ---EAVFIHESYTHDAGFND-DIALIKLNKRVVINSNTPICLPRK-BAESPMFRDDIG 568
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: DB 843 VQVRALAEIHVNPNDPGNLNFDLALIQKTYTLTRVQPLCLPDDITTRRLKSGTTLA 902
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: QY 569 TASGMLTQRCGLANLMLVDPIDVDHOKCTAYAEKPPYPRGSSVTANMLCAGIESGKDS 628
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 903 VVTGGLNENNTYSTIQOAVLPVVAASTCEGKYHEADPL-TVTENNFCAQYKKGRIDA 961
Oy 629 CRDSSGGLVFL-DETE-RMFVGVISWGS-MNCGAQGVYTKVINYIPMIENII 683
Db 962 CSDSGGLPVFADBSRTERRWVLEGVISWGSPPCCGKANQYGTFTKVNFILMIROFI 1019

US-08-596-405-4
; Sequence 4, Application US/08596405
; Patent No. 5858706
; GENERAL INFORMATION:
; APPLICANT: Ding, Jeak Ling
; APPLICANT: Ho, Bow
; TITLE OF INVENTION: The Cloned Factor C cDNA of the
; TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinus scorpilis
; TITLE OF INVENTION: rctundicauda and Purification of Factor C Proenzyme
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,405
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1781-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1019 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-596-405-4

Query Match 12.5%; Score 470.5; DB 2; Length 1019;
Best Local Similarity 28.1%; Pred. No. 2.7e-31;
Matches 151; Conservative 63; Mismatches 187; Indels 137; Gaps 26.

Oy 265 HRIET-----KSNVY---TTFV-VTDSGDHTG-----WKIHNTSTA 297
Db 500 HRLDAKEPFIWELMDRSNVNLNDNTTFWASGEGNETNCVYMDIDQDLSVKT--KSCF 557
Oy 298 QP-----CPYMAEPNGHVSFVQAKYILKDSFSIFCEFGYLLGCHPL 341
Db 558 QPSSFCAMMDLSDRNKAKCDDPGSLGNHAT-LHGOST--DGFYAGSSIRYGCYVLIYLS 614
Oy 342 KSFYAVCOKDGSWDRPMAC-SIVYC-GPPDDLPS-GRVEYITFGCVTT----- 387
Db 615 GTEYTCCTNGTWSAPKRCIKVITCOMP--VPSYGYE-IKPPSRNISIRVGSPPFLR 671
Oy 388 -----YKAVIO--YSCETFTYMKVND 407
Db 672 LPLRLPLPLARAKPPPKRRSSQPTVYDLASKYKLEBGRHVRVSSRAIYTCESRYTELGSQ 731
Oy 408 GKYVEADGEFWTSSKGEKSLPYCEPVCGLISARTTGRIYGGOKAKPGDFPMOVL----- 462
Db 732 GGR-CDSGNMGSGRPAS-----CIPVCGRSDSPRSPFIWNGNSTELGQPMQAGISRWLA 785

```


US-08-296-014A-2

Query Match	12.5%;	Score 470.5;	DB 1;	Length 1083;
Best Local Similarity	28.1%;	Pred. No. 2.9e-31;		
Matches 151; Conservative	63;	Mismatches 187;	Indels 137;	Gaps 26

```

QY 265 HRIET-----KSNV---TTF-VTDESGHGT-----WKIHVSTA 297
Db 564 HRDAEKPEFIWELMDSNVYLNINLTFAASGEGENETNCVYMDIODOLQSVKT--KCF 621
QY 298 QP-----CPYPAPNGHVSPVQAKYILKDSFSIFCETGYELLQHLPL 341
Db 622 QPSSFACAMDLSRNNKAKCDPPGSLENGHAT-LHGSI--DGFYAGSSIRYSCVLIHLS 678
QY 342 KSPFAYCQKGDSDRPMAPAC-SIVDC-GPPDDLPS-GREYITGAGVTT-----387
Db 679 GTETVYCTTGTSARPKRCIKYITCONPP--VPISGSYE-IKPSRTNISRVGSPFLR 735
Db 388 -----KVAIYO--YSCSETFYIMKYND 407
QY 736 LPRLLPLARAARPPKPRSSQPSIVDLASKVYLPFGHYRVGSRMIYTCESRYELLSQ 795
Db 408 GKXYCEADGFWTSSKGEKSLPCEPCEGSLARTGRTIGGAKARGDEPMQVYL-----462
QY 796 GRR-CDSSNGMWSGRPAS-----CIPVCGKSDSPREFIWMNGSTEGQPMWAGISRWLA 849
Db 463 ---LGTTAAGALLYDWMVLTAAHAYEEDKHDSALDIMGTLKRLSPHYTQAWG-----514
QY 850 DHNMWFLQCGGSLINKEKMITAHCV--TYSATLIDPNQFKMYLKYVHDSRDDY 906
Db 515 ----EAFIHGEGTHDAGFON-DIALIKLNKKVYINSNTITPCLEPRK-EAESFMATDIDG 568
QY 907 VQVEALEIHYVNPYDGNLFDIALIOKTPVTLITTRQPTCLPDDITTRHLEEGTLA 966
Db 569 TASCWGCTQCGFLARNLMYVDIIVDHOKTAAEYERPPRGSVYANMLCAGLESKGDS 628
QY 967 VVTGMLGNENNTYSETIQQVLPVVAASICEBYKRADELPL-TVLENMFCAGYKKGRDA 1025
Db 629 CRRGSGCALVFL-DSLEF-RWFYGGIYVSWG--MNGCEAGQCYVYIKVINYIPIWENIT 683
QY 1026 CSGSGGSLVFLFADDSRTERRWVLEGTIVSWGSPGCKRAKQAYGGKFLVNVVFLSMIOFI 1083

```

: Sequence 2, Application US/08596405
 : Patent No. 5858706
 : SERIAL INFORMATION:
 : APPLICANT: Ding, Jeak Ling
 : APPLICANT: Ho, Bow
 : TITLE OF INVENTION: The Cloned Factor C cDNA of the
 : TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus
 : TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzym
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Birch, Stewart, Kolasch & Birch
 : STREET: 8110 Gatehouse Road, Suite 500 East
 : CITY: Falls Church
 : STATE: Virginia
 : COUNTRY: USA
 : ZIP: 22042
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/596,405
 : FILING DATE:
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Murphy, Jr., Gerald M.
 : REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105F

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
FAX: (703) 205-8000

INFORMATION FOR SEQ ID NO: 2
SEQUENCE CHARACTERISTICS:

MOLECULE TYPE: protein
US-08-596-405-2

Query Match	12.5%;	Score 470.5;	DB 2;	Length 1083;
Best Local Similarity	28.1%;	Pred. No. 2.9e-31;		
Matches 151;	Conservative 63;	Mismatches 187;	Indels 137;	Gaps 26

[illegible]

```

: RESULT 8
: US-08-877-620-2
: Sequence 2, Application US/08877620
: Patent No. 5985590
: GENERAL INFORMATION:
: APPLICANT: Ding, Jeak Ling
: APPLICANT: Ho, Bow
: TITLE OF INVENTION: The Cloned Factor C cDNA of the
: TITLE OF INVENTION: Singapore Horseshoe Crab, Carinoscorpis
: TITLE OF INVENTION: rotundicauda and Purification Of Factor C Proenzyme
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: 810 Gatehouse Road, Suite 500 East
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22042
: COMPUTER READABLE FORM:
:

```


Db 252 ADVSLSTFRSFVAPCDHEGSDLVTVYDLSLSPMEHVAVRLCG----- 295
QY 108 SLGSLDITFRSDY-----SNEKPTGFPEA-FYAEDIDECQVAPGAPTCDDHCH 157
Db 296 TFPSPYNLFLSSQNFVLVTLINTGRHGFETFPQLPKMSSC----- 340
QY 158 NHUGFYCSCRAGYVLRHNRKTCGALSCGOVFTQSRGELSSPEYPRPYPLSSCTYSISL 217
Db 341 -----GGVLSDTQGTFSPPYCPHYPNINCTWMIKV 372
QY 218 EBFVSYLDFVESFDVETH-PETLCYDELKIDTDRREHPCFGKTLPHLEKSNVTI 276
Db 373 PNRNNAVKRKLFLYLDPNVPVSGCTKDYVEINKEKGS-----GERSQFVSSNSKITY 427
QY 277 TFFVDESQDTGKMIHTS--TAOPCPYPAPNGHVSPOAK-----YILKDSFS 325
Db 428 HFHSDBSYTDTGLAELSLSDSNPCPGMCKTCRIRKELRCDCWADCPDI--SDERY 485
326 IFCEETGYELLQHLPLKSTFATVQK-----DGSMDR--PMPA---CSIYDCGPPDDLPS 374
486 CRCNATHQFCCKNQFCPLFWVCDSDVNDGSDGDEGCSPAGSFKCSNGKCLPQSQKN 545
QY 375 GRVEYITGPVTTYKAVIOVSCSEETPYTMKVNDGKYVC-----EADGFYTSKG--EKS 426
Db 546 GKDNCGDSDSEASCDVNVVSC--TKYTYRCQNG--LCLSKGNPECDCKDCSDGSDEN 601
QY 427 LPYCEPYCGLSARTTGGRIGYGOKAKPGDFPMQVLI--LG-GTTAAGALLYDMVLTAAH 483
Db 602 ---CD--CGLRSTTKQARVYGGTNADGEMPMOVSLLHALQGLHLCASLISPMVLASAH 656
QY 484 AVEEQKH---DASALDIRMGL--KRLSPHYQAMSEAVFHEGYTHDAGFNDIALI 536
Db 657 CFODDNKFKSYDITMTAFGLIDQSKRSASGVQELKRLIITHPSF-NDFTDYDIALI 725
QY 537 KLNKKVYNSNIPICLPKREAESFMRTDIDIGASGMLQRGFL-ARNLMYDIPYVDH 555
Db 716 ELKSYESTSVVRPICLP--DATHVPPAGAIWVTGWHKEGTGLLLOKCEIRINO 773
QY 596 OKCTAAVEKPPRGSVTANMLCAGLESQKDCRCDGSGALVFLDSETRMFWGVGIVSW 655
Db 774 TTCEEDLMPQ-----QITPRMNCVGLSGVDSCQGDGSGPLSSAEND-GRMQAGVYSW 826
QY 656 GSNMCGEAGQYVYTKV 672
Db 827 GE-GCAQRNKPQVYTRL 842

11
944-483-29
Sequence 29, Application US/08944483
Patent No. 6232456
GENERAL INFORMATION:
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: KLASSEL, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6232456e
US-08-944-483-29

Query Match 11.5%; Score 430.5; DB 4; Length 242;
Best Local Similarity 39.0%; Pred. No. 8.7e-29;
Matches 98; Conservative 41; Mismatches 81; Indels 31; Gaps 10;

QY 445 IYGGOKAKPGDFPMOYLIIGTTAAGALLYDNVLTAAHVAVEQKHDA---SALDIRMG- 500
Db 1 IIGGOKAKNGNFPWQFTNHHGCGALLDKNWILTAHTLYKEHEAOSNADLVFLGH 60
QY 501 ---TKRLSPHYTQAMSEAVFHEGYTHDA--GPDNDIALIKLNKKVYNSNIPICLP 554
Db 61 TNYEELMKLGNHPIR---RVSVPDYRODESYNFEGLDIALLELENSVTLGPULPICLP 116
QY 555 RKAESFMRTDIDIGTASQWLTQRGFLARLMTVVDIPIYDHOKCTAAVEKPPRPS--- 611
Db 117 -DNDTFYDGLMGVYSGFVMEK-IAHDLRFVRLPVANPOAC-----ENMLRKKNRM 167
QY 612 --VTANMLCAGLESQKDCRCDGSGALVFLDSETRMFWGVGIVSGSNMCGEAGQYV 669
Db 168 DVFSQNMFCAGHPSLKQDACQDSCGVFAVRDPNTDRWATGIVSWG-IGCSNG--YGFY 224
QY 670 TKVINYIPWIE 680
Db 225 TKVINYVDMIK 235

RESULT 12
US-08-200-900A-2
Sequence 2, Application US/08200900A
Patent No. 5655566
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

Db 454 -ARCVRLENGTDS--SGLVQF-----RIGSIMHVAEEN--WTQISDD--VCOLLGL 500
QY 418 WTSSKRGKSLPVE-----PYCGLSARTTG-----443
Db 501 GT--GSSVPTSTGGGPPVNTAPNGLILTPSOQLEDLILLQCNKSGCKLV 557
QY 444 -----RIVGQKAKPGDFPMQV-LILGTTAAGA-LLYDMNVLTAAHAYEOKHDAS---493
Db 558 QEVSPKIVGSGDSREGMPVVALYPDDQVCCASLVSRLMVASAHCVGRNNEPKWK 617
QY 494 -ALDIRMGLKRLSPHYTQAMSEAVFIEGTYTDAGFDNDIALIKLNKVVINSITPIC 552
Db 618 AVLGLHMAS-NTSPQIETRLIDQIVINPHY-NKRRKNNDIAMHLEMKVNTDYIQPIC 675
QY 553 LPRKEASEMRTDDIGASWAG-LTORGFLARNLMYVDIPVHOKCTAAVEKPPYRGS 611
Db 676 LP--EENQVPPERICSIACMGALIVOGSTADVLQEADVPLLSNEKC-----QQMPEYN 728
612 VTANMLCAGLESCKDSCRGDSGALVFLDSETERMFVGIVSGMNCGEAGQGYTK 671
729 ITENMVCAGIEAGGVDSGCCDSGGL--MCOENNRMLLAGVTSFG-YQCALPNRPGVIAR 785
QY 672 VINIPIENIT 683
Db 786 VPRFTWIIQSFL 797
RESULT 14
US-08-295-411-3
; Sequence 3, Application US/08295411
; Patent No. 5679639
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H.
; APPLICANT: Mesters, Rolf M.
; TITLE OF INVENTION: Serine Protease-Derived Polypeptides and
; TITLE OF INVENTION: Anti-Peptide Antibodies, Systems and Therapeutic Methods
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Office of Patent Counsel, The Scripps
; ADDRESSEE: Research Institute
; STREET: 10666 No. 5679639th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,411
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,989
; FILING DATE: 18-NOV-1991
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Filling, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI263.0C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 448 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..139
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: Region
; LOCATION: 140..142
; OTHER INFORMATION: /note= "Factor X Connecting
; OTHER INFORMATION: tripeptide"
; NAME/KEY: Region
; LOCATION: 143..448
; OTHER INFORMATION: /note= "Factor X Heavy Chain"
US-08-295-411-3
Query Match 10.8%; Score 405; DB 1; Length 448;
Best Local Similarity 23.6%; Pred. No. 3e-26;
Matches 142; Conservative 81; Mismatches 173; Indels 206; Gaps 28;
QY 90 CGOESTTERAPGKDFYSLGSSLDITFRSDYSNEKPFTEPAFYAEDIDECOVAPGEA 149
Db 17 CMEETCSYEBA--REVFE-----SDKTNE-----FWNKY--KDDQDETSP---54
QY 150 PTCDDH--CHNHLGGEYCSGRAGYVLRNKRKTCALCSGGVFORSGELSSPEVPRPYK 207
Db 55 --CONQCKRBGLGEYTCLEGF---EGRNC-----ELFTRKLCSLDN-----93
QY 208 LSSCTYSISLEEGFVILDFEVESFDVETHPETLCPPYDFLKIQTDRBHGPF-CGK-TLPH 265
Db 94 -GPCD-QFCHEDQNSVYSCARGYTLADNGKACP-----TGYPGCKQTLER 139
QY 266 RIETKSNVTITTYVTDSCGHTGKIHVSTADPCPTMAPRPNHGSPVQAKIYLDSEFS 325
Db 140 R--KRSVAQATSSSGEAPDSITWK-----PYDAA-----DLDPTEPFLLD---179
QY 326 IFCEGYELLQGHLPKLSFTAVQCKGSMWRPMPACISVDCGPPDDLPSGRVEIITRGV 385
Db 180 -FNOTOPE-----186
QY 386 TTYKAVIQSCETFTYTMKYNQKGVCEADGFWTSKGEKSLPVEPCGLSARTTGRI 445
Db 187 -----RGDNNTL-----RI 195
QY 446 YGQKAKPGDFPMQVILGTT--AAGALLYDMNVLTAAHAYEOKHDASALDIRMGL 502
Db 196 VGGQCKKDCGCPMQUALLINEENEGFCGTTLSEFYILTAHCLYQAKR---FKVRVGD 251
QY 503 KRISPHYTOAMS--EAVFIEGTYTHDAGFPNDIALIKLNKVVINSITPICPRKE-AE 559
Db 252 NTDOEBSGEAVHEVEYVVKINRFTKEP-YPDFIALVLRKPIYFRMVAACLPERRMAE 310
QY 560 SFMTDDIGTASGMLT-QRGFLARNLMYVDIPVHOKCTAAVEKPPYRGSVTANMLC 618
Db 311 STLMGKTGIVSGGRHHEGROSTRKMLEVYPVVDNSKLS-----SSFIITQMF 364
QY 619 AGLESCKDSCRGDSGALVFLDSETERMFVGIVSGMNCGEAGQGYTKVINYIPW 678
Db 365 AGYDTKOEDACOGDSGPGPHYTRFKDT--YFVTGIVSWGE-GCARKKGYYITVTATLKN 421
QY 679 IE 680
Db 422 ID 423
RESULT 15
US-08-955-471-3
; Sequence 3, Application US/08955471
; Patent No. 5968751
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H.
; APPLICANT: Mesters, Rolf M.

APPLICANT: Solomon, Robert A.
APPLICANT: Schmidt, Ann Marie
TITLE OF INVENTION: METHODS FOR TREATING AN ISCHEMIC DISORDER AND IMPROVING
TITLE OF INVENTION: STROKE OUTCOME
FILE REFERENCE: 51917-B
CURRENT APPLICATION NUMBER: US/09/053,871A
CURRENT FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 461
TYPE: PRT
ORGANISM: Homo Sapien
US-09-053-871A-21

Query Match 10.8%; Score 405; DB 4; Length 461;
Best Local Similarity 30.5%; Pred. No. 3.1e-26;
Matches 107; Conservative 50; Mismatches 92; Indels 102; Gaps 16;

396 CEETFTYMKVNDGKYVCEADGFWTSKGEKSLPVCPEV---CG-LSARTTG----- 442
141 CEO--FCKNSADNKVYVCSCTEGRLAENOKS---CEPAVFPFGGRVSVSOTSKLTRAETV 195
443 -----GRYGGOKAKPDPEPMQVLLIGTTA--AGA 471
196 FPDVYVNSTEAEITLDNITOSTQSFNDFTRVYGGEDAKGQFPQVVLNGKVDAFCGGS 255
472 LLYDNVLTAAHAY-----YEOKHNASALDIRMGLKRLSPHYTQ 511
256 IVEKMWLVTAHCEVETGKITYVAGEHNIETEETEQKRN-----VIRIIPHN- 304
512 AMSEAVFIEHGYTHDAGFNDIALIKLNKVINSNTPICLPKREASEPMRTDICTAS 571
305 -YNAAT--NKYNH-----DIALLEDEPLVLSYVTPICIAKDEYTNIFLFGSGYVS 354
572 GWC-LTORGFLARNLMYVDIPVDHOKCTAAEKKPPYPRGSVTANMLCAGLESGRDSCR 630
355 GWCGRVHFHGRSALVQLRVPLVDRAFCRS-----TKFTIYNNMFCAGFHGGRDSCQ 408
631 GDSGGALVPLDSETE-RMFVGGIVSNGSMNGEAGQVGYTKYINITYPTE 680
409 GDSGGPHV--TEVEGTSFLTGLISWGE-ECAMKKGKGYITKVSRYVNMWK 455

RESULT 19
5521070-2
Patent No. 5521070
APPLICANT: MEULIEN, PIERRE
TITLE OF INVENTION: DNA SEQUENCE CODING FOR HUMAN FACTOR
A SIMILAR PROTEIN, EXPRESSION VECTOR, TRANSFORMED CELLS,
AND FOR PREPARING FACTOR IX AND CORRESPONDING PRODUCTS OBTAINED
NUMBER OF SEQUENCES: 6
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,489
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 970,966
FILING DATE: 03-NOV-1992
APPLICATION NUMBER: 433,276
FILING DATE: 08-NOV-1989
SEQ ID NO: 2;
LENGTH: 461
5521070-2

Query Match 10.8%; Score 405; DB 6; Length 461;
Best Local Similarity 30.5%; Pred. No. 3.1e-26;
Matches 107; Conservative 50; Mismatches 92; Indels 102; Gaps 16;
396 CEETFTYMKVNDGKYVCEADGFWTSKGEKSLPVCPEV---CG-LSARTTG----- 442
141 CEO--FCKNSADNKVYVCSCTEGRLAENOKS---CEPAVFPFGGRVSVSOTSKLTRAETV 195
443 -----GRYGGOKAKPDPEPMQVLLIGTTA--AGA 471

196 FPDVYVNSTEAEITLDNITOSTQSFNDFTRVYGGEDAKGQFPQVVLNGKVDAFCGGS 255
472 LLYDNVLTAAHAY-----YEOKHNASALDIRMGLKRLSPHYTQ 511
256 IVEKMWLVTAHCEVETGKITYVAGEHNIETEETEQKRN-----VIRIIPHN- 304
512 AMSEAVFIEHGYTHDAGFNDIALIKLNKVINSNTPICLPKREASEPMRTDICTAS 571
305 -YNAAT--NKYNH-----DIALLEDEPLVLSYVTPICIAKDEYTNIFLFGSGYVS 354
572 GWC-LTORGFLARNLMYVDIPVDHOKCTAAEKKPPYPRGSVTANMLCAGLESGRDSCR 630
355 GWCGRVHFHGRSALVQLRVPLVDRAFCRS-----TKFTIYNNMFCAGFHGGRDSCQ 408
631 GDSGGALVPLDSETE-RMFVGGIVSNGSMNGEAGQVGYTKYINITYPTE 680
409 GDSGGPHV--TEVEGTSFLTGLISWGE-ECAMKKGKGYITKVSRYVNMWK 455

RESULT 20
US-08-073-531B-1
Sequence 1, Application US/08073531B
Patent No. 5621039
GENERAL INFORMATION:
APPLICANT: Hallahan, et al.
TITLE OF INVENTION: Factor IX - Polymeric Conjugates
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: GALGANO & BURKE
STREET: 300 Radro Drive
CITY: Hauppauge
STATE: New York
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS 6.0
SOFTWARE: Wordperfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,531B
FILING DATE: June 8, 1993
CLASSIFICATION: 525
PRIOR APPLICATION DATA:
APPLICATION NUMBER: -
FILING DATE: -
ATTORNEY/AGENT INFORMATION:
NAME: GALGANO & BURKE
REGISTRATION NUMBER: 30,735
REFERENCE/DOCKET NUMBER: 128-7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 582-6161
TELEFAX: (516) 582-6191
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
MOLECULE TYPE: Unknown to applicant
HYPOTHETICAL: -
ANTI-SENSE: -
ORIGINAL SOURCE:
ORGANISM: -
INDIVIDUAL ISOLATE: -
CELL TYPE: -
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:

JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 415.

Query Match 10.8%; Score 404; DB 1; Length 415;
Best Local Similarity 30.8%; Pred. No. 3.3e-26;
Matches 108; Conservative 49; Mismatches 92; Indels 102; Gaps 16;

QY 396 CEETFTYMKVNDKRYVCEADGFWTSKGEKSLPYCEPV-----CG----- 435
DB 95 CEQ--FCKNSADKVKVCSCEGYRLAENQKS---CEPAVFPFCGRVSVSQTSLTRAFAV 149
QY 436 -----LSARTTG-----GRYGGOKAKRGDFPMOYLLIGTTA--AGA 471
150 FPDVYVNPTEAETILDNITQGTSEFNDFTRVYGGEDAKGQFPWOYVLNGKYDAFCGGS 209
472 LLYDNWVLTAAHAY-----YEOKHDSALDIMGTLKRLSPHYTQ 511
DB 210 IYNEKMTVTAHCHVEYGVKITTVVAGEHINETETEDKRN-----VIRIIPHNN- 258
QY 512 AMSEAVFIHEGYTHDAGFNDIALIKLNKVVINSNITPICLPKREAESPMRTDICTAS 571
DB 259 -YNAAI--NKYNN-----DIALLEDEPLVLSYVTPICIDAKKEYTNIFLKFGSGYVS 308
QY 572 GNG-LTORGFLARKMLYVDIPIYDHOKCTAAYEKPPYPRGSVTANMICALGLESCKDSCR 630
DB 309 GNGRVFHKGRSALVLELYLRVPLVDRAICLRS-----TKFTIYNNMFCAGFHGGRDSCQ 362
QY 631 GDSGALVFLDSETE-RMFVGGIVSMGSMNGCEAGOGYVTKVINYIPMIE 680
DB 363 GDSGGRPHV---TEVEGTSFLTGIIISMG-ECAMKKGKIGITTKVSRYVNMWK 409

RESULT 21

US-08-766-288-1
Sequence 1, Application US/08766288
Patent No. 5969040

GENERAL INFORMATION:
APPLICANT: Hallahan, et al.
TITLE OF INVENTION: Factor IX - Polymeric Conjugates
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: GALGANO & BURKE
STREET: 300 Radro Drive
CITY: Hauppauge
STATE: New York
COUNTRY: USA
ZIP: 11788

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS 6.0
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,288
FILING DATE: -
CLASSIFICATION: 525
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,531
FILING DATE: June 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: GALGANO & BURKE
REGISTRATION NUMBER: 30,735
REFERENCE/DOCKET NUMBER: 128-7 (DIV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 582-6161
TELEFAX: (516) 582-6191
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 415 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown to applicant
MOLECULE TYPE: -
HYPOTHETICAL: -
ANTI-SENSE: -
ORIGINAL SOURCE:
ORGANISM: -
INDIVIDUAL ISOLATE: -
CELL TYPE: -
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
PUBLICATION INFORMATION:
AUTHORS:
TITLE: JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 415.

Query Match 10.8%; Score 404; DB 2; Length 415;
Best Local Similarity 30.8%; Pred. No. 3.3e-26;
Matches 108; Conservative 49; Mismatches 92; Indels 102; Gaps 16;

QY 396 CEETFTYMKVNDKRYVCEADGFWTSKGEKSLPYCEPV-----CG----- 435
DB 95 CEQ--FCKNSADKVKVCSCEGYRLAENQKS---CEPAVFPFCGRVSVSQTSLTRAFAV 149
QY 436 -----LSARTTG-----GRYGGOKAKRGDFPMOYLLIGTTA--AGA 471
150 FPDVYVNPTEAETILDNITQGTSEFNDFTRVYGGEDAKGQFPWOYVLNGKYDAFCGGS 209
472 LLYDNWVLTAAHAY-----YEOKHDSALDIMGTLKRLSPHYTQ 511
DB 210 IYNEKMTVTAHCHVEYGVKITTVVAGEHINETETEDKRN-----VIRIIPHNN- 258
QY 512 AMSEAVFIHEGYTHDAGFNDIALIKLNKVVINSNITPICLPKREAESPMRTDICTAS 571
DB 259 -YNAAI--NKYNN-----DIALLEDEPLVLSYVTPICIDAKKEYTNIFLKFGSGYVS 308
QY 572 GNG-LTORGFLARKMLYVDIPIYDHOKCTAAYEKPPYPRGSVTANMICALGLESCKDSCR 630
DB 309 GNGRVFHKGRSALVLELYLRVPLVDRAICLRS-----TKFTIYNNMFCAGFHGGRDSCQ 362
QY 631 GDSGALVFLDSETE-RMFVGGIVSMGSMNGCEAGOGYVTKVINYIPMIE 680
DB 363 GDSGGRPHV---TEVEGTSFLTGIIISMG-ECAMKKGKIGITTKVSRYVNMWK 409

RESULT 22

PCT-US92-10068-1
Sequence 1, Application PC/TUS9210068

GENERAL INFORMATION:
APPLICANT: Altieri, Dario C
APPLICANT: Edgington, Thomas S
TITLE OF INVENTION: Factor X-Derived polypeptides and
TITLE OF INVENTION: Anti-Peptide Antibodies, Systems and Therapeutic Methods
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Office of Patent Counsel, The Scripps
ADDRESS: Research Institute
STREET: 10666 North Torrey Pines Road
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

Query Match	10.7%	Score 401	DB 2	Length 1013
Best Local Similarity	28.2%	Pred. No. 2.1e-25		
Matches 120	Conservative 65	Indels 90	Gaps 19	
QY 11 CGSVATPLGPKWPEVFGRLASPFPGCYANDQERWTLTAPPGYRLRYFTFHEDLEISH 70				
11	1	1	1	1
11	1	1	1	1

1 CLASSIFICATION: 2
3 PRIOR APPLICATION DATA: 4
5 APPLICATION NUMBER: 60/034,471 6
7 FILING DATE: 02-JAN-1997 8
9 ATTORNEY/AGENT INFORMATION: 10
11 NAME: PRESTIA, PAUL F 12
13 REGISTRATION NUMBER: 23,031 14
15 REFERENCE/DOCKET NUMBER: ATG-50038 16
17 TELECOMMUNICATION INFORMATION: 18
19 TELEPHONE: 610-407-0700 20
21 TELEFAX: 610-407-0701 22
23 TELEX: 846169 24
25 INFORMATION FOR SEQ ID NO: 2: 26
27 SEQUENCE CHARACTERISTICS: 28
29 LENGTH: 1013 amino acids 30

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
us-08-991-408-2

Query Match 10.7%; Score 401; DB 3; Length 1013;
Best local similarity 28.2%; Pred. No. 2.1e-25;
Matches 120; Conservative 65; Mismatches 151; Indels 90; Gaps 19;

QY 11 CGSVAPPLGKWPPEVGRGLASGPFGEYANDQERRWTLTPPGYRLRLFTHDLFLSH 70
DB 618 CGGLTKRLN-----GRTTPGWPKEYPNNKNCWQVAPTYQYRISVKFEFELEGNE 669
QY 71 LCEYDFVK---LSSGAKVLATLCGOESTDTERAPGKDTFYSLGSSLDITFRSDYSNEKP 126
DB 670 VCKYDYVEITWSGLSSEKSLHGKFCGAEV-----PEVITSGFNNNRITFKSDNTYSK- 720
QY 127 FTGFEEAFYAEDIDECQVAPGEAPTCDHGCHNHGGEFYCSCRAGYVLHRRKRTCSALCSG 186
721 -KGFKAHFFS-DKDECSKONG---GCOHECVNTMGSYMCCCRNGFVLHDKKHCKEACE 775
187 QVETQSGELSSPEYPRPYKLSCTYSISLEEGFSYILDVSESDEVETHPETLCPYDFL 246
776 QKIHSPBGLITSPNWPDKYPSREKCTWEISATPGHRIKIAFSE-FELEQHOE--CAYDHL 832
QY 247 KIQTDRREHGP---FCGKTLPHRIETKSNVTITFEVTDSDGDTGWKIHYS-----T 296
DB 833 EVFDGETEKSPILGRICGNKIPDPVATGKNMFRFVSDASVQKGFQATHSTECGRK 892
QY 297 AQPCP-----YPMAPPNGHVPVQAKYIKD-----SFSIF-----CETGYELLQGH 338
DB 893 AESKPRDLYSHAQFGDNNYPGQVDCLEWLVSEKSRLELSFQTFEVEEADCGDYVELF 952
QY 339 LPLKSFYAV-----CQKDGSMDRMPACSLVDGPPDDLPSGRVEYITGPGVTTYKAVIQ 393
DB 953 DGLDS--TAVGLGRFC--GS-----GPREEL-----YSIGDSV-----LIH 984
QY 394 YSCDET 399
DB 985 FHFDIT 990

Search completed: January 23, 2003, 22:20:30
Job time : 28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 23, 2003, 22:15:51 : Search time 52 Seconds
(without alignments)
1266.235 Million cell updates/sec

Title: US-09-874-198-2

Perfect score: 3758

Sequence: 1 MRLRLTLGLCGSVATPLGP.....GYVTKYINYPMENTISDF 686

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

hed: 283224 segs, 96134422 residues

number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3745	99.7	686	1 A59271	Ra-reactive factor
2	1540	41.0	699	1 I54763	Ra-reactive factor
3	1412	37.6	705	1 C1H0RB	complement subcomp
4	1244.5	33.1	695	1 S05008	complement subcomp
5	1238	32.9	688	1 C1H0S	complement subcomp
6	1214	32.3	694	2 JC6554	complement subcomp
7	541.5	14.4	1524	2 JC6037	polypeptide - Afri
8	471.5	12.5	1019	2 A38738	coagulation factor
9	467	12.4	1004	2 T30338	coagulation factor
10	466.5	12.4	855	2 JC7731	membrane-bound arg
11	466.5	12.4	855	2 JC7775	membrane type-seri
12	443	11.8	1019	1 A56318	enteropeptidase (E
13	442.5	11.8	400	1 A48050	coagulation factor
14	424.5	11.3	1464	2 S58984	development protei
15	422	11.2	475	1 EXCH	coagulation factor
16	418.5	11.1	1034	1 A53663	coagulation factor
17	416.5	11.1	991	2 I49540	enteropeptidase (E
18	414.5	11.0	986	1 B58788	procollagen C-endo
19	412	11.0	452	1 A30351	coagulation factor
20	408.5	10.9	1035	1 A43090	enteropeptidase (E
21	406.5	10.8	1070	2 T31069	coagulation factor
22	405	10.8	461	1 KFH0	coagulation factor
23	405	10.8	618	2 A35827	thrombin (EC 3.4.2
24	403	10.7	329	1 HPDG	thrombin (EC 3.4.2
25	403	10.7	488	1 EXHU	coagulation factor
26	402.5	10.7	285	1 I48144	coagulation factor
27	402	10.7	347	2 G00006	haptoglobin - blac
28	401.5	10.7	416	1 S33777	hepsin (EC 3.4.21
29	400.5	10.7	459	2 J00419	coagulation factor

30	399.5	10.6	406	1 HPHU2	haptoglobin precur
31	399.5	10.6	416	1 KFB0	coagulation factor
32	397	10.6	461	1 S18994	protein C (activat
33	394	10.5	347	1 HPMS	haptoglobin precur
34	393.5	10.5	271	2 I46580	factor IX - p19 (f
35	391	10.4	347	1 HPRT	haptoglobin precur
36	389.5	10.4	282	2 I84621	coagulation factor
37	387	10.3	347	1 HPHU1	haptoglobin precur
38	387	10.3	625	1 TBBO	thrombin (EC 3.4.2
39	387	10.3	730	1 BMHU1	procollagen C-endo
40	385	10.2	348	1 HPHUR	haptoglobin-relate
41	384	10.2	461	1 JX0210	protein C (activat
42	383.5	10.2	275	2 I46712	factor IX - rabbit
43	383	10.2	345	2 I36941	haptoglobin - chim
44	382	10.2	461	1 KKHU	haptoglobin - chim
45	382	10.2	622	1 TBHU	thrombin (EC 3.4.2
46	381.5	10.2	417	1 S00845	hepsin (EC 3.4.21
47	381.5	10.2	823	1 A58788	procollagen C-endo
48	380	10.1	349	2 I36944	haptoglobin - chim
49	378	10.1	251	2 PC1235	29K serine protein
50	378	10.1	482	1 EXRT	coagulation factor
51	377	10.0	346	2 I36942	haptoglobin - chim
52	376.5	10.0	456	1 KXBO	protein C (activat
53	376.5	10.0	638	1 K0MSPL	plasma kallikrein
54	374	10.0	1057	1 A39288	dorsal-ventral pat
55	374	10.0	443	2 I46932	coagulation factor
56	369	9.8	492	1 EXBO	coagulation factor
57	368.5	9.8	810	2 B30848	plasma (EC 3.4.21
58	367.5	9.8	407	1 KFB07	coagulation factor
59	367	9.8	259	2 S68424	allergen Der f III
60	367	9.8	638	1 KOHUP	plasma kallikrein
61	362	9.6	343	1 A57014	protrastin (EC 3.4
62	361	9.6	466	1 KFHU7	coagulation factor
63	360.5	9.6	617	2 S10511	thrombin (EC 3.4.2
64	360	9.6	790	1 PLPG	plasma (EC 3.4.21
65	359.5	9.6	250	2 S55493	serine proteinase
66	359.5	9.6	810	1 PLHU	plasma (EC 3.4.21
67	358	9.5	767	2 T30018	hypothetical prote
68	357	9.5	812	1 PLBO	plasma (EC 3.4.21
69	356.5	9.5	638	1 K0RTPL	plasma kallikrein
70	355.5	9.5	274	2 I47078	coagulation factor
71	352.5	9.4	237	2 S55378	serine proteinase
72	352.5	9.4	707	2 JC2218	procollagen C-endo
73	352	9.4	812	1 PLMS	plasma (EC 3.4.21
74	351	9.3	375	1 A23689	limulus clotting e
75	349	9.3	247	2 S05494	trypsin (EC 3.4.21

ALIGNMENTS

RESULT 1
Ra-reactive factor (EC 3.4.21.-) precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence, revision 19-May-2000 #text, change 16-Jun-2000
C:Accession: A59271
R:Thiel, S.; Vorup-Jensen T.; Stover, C.M.; Schwaeble, W.J.; Laursen, S.B.; Poulsen, Nature 386, 506-510, 1997
A:Title: A second serine protease associated with mannan-binding lectin that activate
A:Reference number: A59271, MUID:97242412, PMID:9087411
A:Accession: A59271
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-686 <JEN>
A:Cross-references: GB:Y09926; NID:q4007626; PIDN:CAA71059.1; PID:q4007627
A:Experimental source: tissue: liver
A:Note: submitted to GenBank, December 1996
A:Note: parts of this sequence, including the amino end of the mature protein, were d
C:Genetics:
A:Gene: GDB:MASP2
A:Cross-references: GDB:6071500

A:Map position: 1p36.2-1p36.3
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydroxylase; serine P
E:1-15/Domain: signal sequence #status predicted <SIG>
E:16-444,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>
E:19-134/Domain: C1r/C1s repeat homology <C1r1>
E:142-180/Domain: EGF homology <EGF>
E:184-293/Domain: C1r/C1s repeat homology <C1r2>
E:300-361/Domain: complement factor H repeat homology <FH1>
E:366-430/Domain: complement factor H repeat homology <FH2>
E:445-679/Domain: trypsin homology <TRY>
E:72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,396-430,434-552,
F:158/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:444,445/Cleavage site: Arg-Ile (autolytic) #status predicted
F:483,532,653/Active site: His, Asp, Ser #status predicted

Query Match	99.7%;	Score 3745;	DB 1;	length 686;
Best Local Similarity	99.7%;	Pred. No. 2.8e-241;		
Matches 684;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0;

QY	1	MLLTLTLLGLLGSVAPRLPGKWPPEVFGRLASDPFGGEVANOERWMLTAPGRYRLX	60
	1	MLLTLTLLGLLGSVATPLPGKWPPEVFGRLASDPFGGEVANOERWMLTAPGRYRLX	60
QY	61	FTHPFLDELSHLCEYDFVFKLSSGAKVLATLCSGSESDITERAPGKDTFFYSLGSSLDITERSD	120
DB	61	FTHPFLDELSHLCEYDFVFKLSSGAKVLATLCSGSESDITERAPGKDTFFYSLGSSLDITERSD	120
QY	121	YSNEKPFTGFEAFYAIEDIDECQVAPGEAPTCDDHCHNHLLGSGFYSCRAGYVLLHRNKRTC	180
DB	121	YSNEKPFTGFEAFYAIEDIDECQVAPGEAPTCDDHCHNHLLGSGFYSCRAGYVLLHRNKRTC	180
QY	181	SALCGGQVFTQSSGELSSSEYRPRYPKLSSCTYSLSEEGFVIIIDFVSPFEVTHPELT	240
DB	181	SALCGGQVFTQSSGELSSSEYRPRYPKLSSCTYSLSEEGFVIIIDFVSPFEVTHPELT	240
QY	241	CPYDLKLTQTDREEHGPFCKTLPRIETKSNVTVTTFEYTDSDGDTGKIHVTSIAOPC	300
DB	241	CPYDLKLTQTDREEHGPFCKTLPRIETKSNVTVTTFEYTDSDGDTGKIHVTSIAHNC	300
QY	301	PYPMAPPNHGHSVPVAKYTLKDSFSTFCETGYELLQGLPLKSFYAVCKDGSMDRMPMA	360
DB	301	PYPMAPPNHGHSVPVAKYTLKDSFSIFCETGYELLQGLPLKSFYAVCKDGSMDRMPMA	360
QY	361	CSIVCGPPDDLPSGRVEITTPGVTYTKAVIYOYSGEEFTYMKKNDGKYVCEADGFWTS	420
DB	361	CSIVCGPPDDLPSGRVEITTPGVTYTKAVIYOYSGEEFTYMKKNDGKYVCEADGFWTS	420
QY	421	SKGEKSLPCEPVCGLSARTTGRIYGGOKAKPGDFPMQVLLIGTAYAGALLYDNWVLT	480
DB	421	SKGEKSLPCEPVCGLSARTTGRIYGGOKAKPGDFPMQVLLIGTAYAGALLYDNWVLT	480
QY	481	AAHAYVEQKHDSALDIRMGITLKRLSPHYTOAMSEAVFTHESYTHDAGFNDIALIKLNN	540
DB	481	AAHAYVEQKHDSALDIRMGITLKRLSPHYTOAMSEAVFTHESYTHDAGFNDIALIKLNN	540
QY	541	KVIVINSNITPICLPKEKAESEFMRDIDIGASGWLQREFLARNLMYVDPIVYDHOKCA	600
DB	541	KVIVINSNITPICLPKEKAESEFMRDIDIGASGWLQREFLARNLMYVDPIVYDHOKCA	600
QY	601	AYEKPPYPRGVSYTNAMLAGLESGGKDCRGDGGALVFLDSETERMPFGGIVSGSNMC	660
DB	601	AYEKPPYPRGVSYTNAMLACGLESGGKDCRGDGGALVFLDSETERMPFGGIVSGSNMC	660
QY	661	GEAGQGYVTKKINIPWIENTISDF	686
DB	661	GEAGQGYVTKKINIPWIENTISDF	686
RESULT 2			
154/763			
N-Reactive factor (EC 3.4.21.-) 1 precursor - human			
N-Alternate names: manose binding protein-associated serine proteinase 1 (MASP-1)			

C.Species: Homo sapiens (man)
C.Date: 19-May-2000 #sequence-revision 19-May-2000 #text-change 16-Jun-2000
C.Accession: J54763; JN0883
R.Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.
Int. Immunol. 6, 665-669, 1994
A.Title: Molecular characterization of a novel serine protease involved in activation
A.Reference number: J54763; MUID:94289349; PMID:8018603
A.Accession: J54763
A.Status: preliminary; translated from GB/EMBL/DDBJ
A.Molecule type: mRNA
A.Residues: 1-699 <SAT>
A.Cross-references: GB:D8593; NID:g790963; PIDN:BAA05928.1; PID:g471128
R:Takada, F.; Takayama, Y.; Hatsuse, H.; Kanakami, M.
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993
A.Title: A new member of the C1s family of complement proteins found in a bactericida
A.Reference number: JN0883; MUID:94059062; PMID:8240317
A.Accession: JN0883
A.Molecule type: mRNA
A.Residues: 1'234,'E',236-284,'G',286-498,'K',500-542,'K',544-642,'S',644-699 <PAK>
A.Cross-references: DDBJ:DI1525; NID:g439712; PIDN:BAA04477.1; PID:g439713
A.Experimental source: liver
A.Comment: This is a serum bactericidal factor that activates complement C4 and C2 co
C.Genetics:
A.Gene: GDB:MASPL; GDB:CRARF; CRARF1; PRSS5; MASP
A.Cross-references: GDB:361104; GDB:330994; OMIM:600521
A.Map position: 3q27-3q28
C.Subfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto
C.Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hy
F:1-17/Domain: signal sequence #status predicted SIG>
F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAR>
F:119-135/Domain: C1r/C1s repeat homology <C1R1>
F:143-181/Domain: Egf homology <EGF>
F:185-294/Domain: C1r/C1s repeat homology <C1R2>
F:301-362/Domain: complement factor H repeat homology <FH1>
F:367-432/Domain: complement factor H repeat homology <FH2>
F:449-691/Domain: trypsin homology <TRY>
E:49,178,407/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:73-91,143-157,153-166,168-185,212,242-260,301-349,329-362,367-414,397-432,436-5
F:153/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:448-449/Cleavage site: Arg-Ile (autolysis) #status predicted
F:490,552/Active site: His, Asp, Ser #status predicted

Query Match	41.0%	Score 1540:	DB 1:	Length 699:
Best Local Similarity	42.2%	Pred. No. 1,1e-94:		
Matches 300:	Conservative 130:	Mismatches 235:	Indels 46:	Gaps 14:
QY	1	MLRLFLGLLGGSVATPLGPKWP--EYVGRGLASGPGGEVANDQERRWTLTAPRGYRL	59	
DB	1	MKMLLYALCPSSKASAHATVEILNNGGQIQSGYFPSPSDEVTWNTVTPGPFRIKL	60	
QY	60	YETHDELDSLCEYDEPKFLSSGAKVLTALCGESTPTERAPGKDFPYSLGSSLDITFRS	119	
DB	61	YFMHNLSSSYLCEYDYKAVETEDQVLATFCGRRTTDEQPGGEVVLVSGSFMSTFRS	120	
QY	120	DYSNEKPTGFEDAFYAEDIDECOVAREAPLTCDHCHNHLGFGYSCRAGYVLIHRNKT	179	
DB	121	DFSNEERTGCDAHMAVDYDECKEREDEELSCDHCHYCHANYIGGYCCRCRGYLIHTDNT	180	
QY	180	CSALCSGQVFORSGEELSSPEYRPRPKLSCSYTSLSEGFSTIIDFVESPDVETPER	239	
DB	181	CRVESDNLVFORGVITSPDFPNRPSSCLVTILEELSEFMNLQEDFIIDQHPREV	240	
QY	240	LCPEYFLKIQTDREHNGFGCKTLPRIETSNVTTFVYDESGDITGKINHSTAPQ	299	
DB	241	PCPDYVYIKIKGPKVLYGFCGEKARPELISQSHSVLLIFHSDNSAENKRGRLSTRAGNE	300	
QY	300	CPYPMAPNGVSHPYQAKYLLKDSFSIFCETGYELLQGHLPKLSFTAVCOKDGSMDRPM	359	
DB	301	CPELDPRVHAKRIEPSQAKYFFKDOVLVSCDYGKVLKDNVMDMDFQIECLKDGTWSKIP	360	
QY	360	ACSTVDCGRPDDLPDSGRREYITTCGCVTTTKAVIOTSCSEELFTYR-KYNDKRYCEADGFW	418	
DB	361	TKCIYDCAPDELHGLITTFSTRNNLTITKYSLEISQDEPKYKLNNTITYCSAGVW	420	


```

Db 474 NFPQVATNINHGREGGALLGDRWILTAHNTLYPKHEHQSNASLDTVLGHTNVEELKLG 533
OY 507 PHYQAMSEAFVHEGYTHDA--GEFNDIALIKLNKRVINSNTPLCPKREASEMRT 564
Db 534 NHPLR-----KVSVPDYRQDESYNEEGDIALELENSVTLGPNLLPCLP--DNDITRYDL 587
OY 565 DDIGTAGSGWLTORGFLARNLMVYDIPVDHOKCTAAVEKPPRGS-----VTANMLCA 619
Db 588 GLMGVSGFGVMEER-IAHDLRFVRLPVANPQAC-----ENWLKGNKRMDFVSGNMECA 640
OY 620 GLESGGDCSCRGDSGALVFLDSETERFVGGIYSGSMNCGEAGQGVYTKVNTYIPWI 679
Db 641 GHPLKQDQACGDSGGVFAVADPNTDRWAVAGIVSMG-IGCSRG--YGFYTKLVNTYDWI 697
OY 680 E 680
Db 698 K 698

```

```

C:Species: Mesocricetus auratus (golden hamster)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
C:Accession: S05008
R:Kinoshita, H.; Sakiyama, H.; Tokunaga, K.; Imajoh-Ohmi, S.; Hamada, Y.; Isono, K.; Sak
FEBS Lett. 250, 411-415, 1989
A:Title: Complete primary structure of a calcium-dependent serine proteinase capable of
A:Reference number: S05008; MUID:89325606; PMID:2753140
A:Accession: S05008
A:Molecule type: mRNA
A:Residues: 1-695 <KIN>

```

```

A:Cross-references: EMBL:X16160; NID:g49621; PIDN:CAA34286.1; PID:g49622
A>Note: part of this sequence, including the amino ends of both the heavy and light chain
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
C:Keywords: beta-hydroxyasparagine; calcium binding; duplication; glycoprotein; hydrolase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:17-133/Domain: C1r/C1s repeat homology <C1R1>
F:22-444/Product: serine proteinase heavy chain #status experimental <HCH>
F:141-177/Domain: EGF homology <EGF>
F:181-293/Domain: C1r/C1s repeat homology <C1R2>
F:300-360/Domain: complement factor H repeat homology <FH1>
F:305-428/Domain: complement factor H repeat homology <FH2>
F:445-682/Domain: trypsin homology <TRY>
F:446-695/Product: serine proteinase light chain #status experimental <LCH>
F:71-89,141-153,149-162,164-177,181-208,240-257,300-347,327-360,365-410,392-428,432-556,
F:555/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:413/Binding site: carboxylate (Asn) (covalent) #status predicted
F:536,638/Active site: His, Asp, Ser #status predicted

```

```

Query Match 33.1%; Score 1244.5; DB 1; Length 695;
Best Local Similarity 39.9%; Pred. No. 4.8e-75;
Matches 271; Conservative 96; Mismatches 269; Indels 41; Gaps 18;

```

```

OY 24 EP-VFGLASPGFGEYANDQERRMTLPARGYRLRYTHFDELSHLCEYDFVKLSG 82
Db 22 EPTMHGELISPNYQAPVAPNEMEKTWIDIEVEGFCVRLYTHLDELSENCEYDVSQISG 81
OY 83 AKVLTATCGQESTTEAPRGKDTFTYSLGSSLDITFRSDYSNEKFTFEAFYAEDIDEC 142
Db 82 GVEBGRICGQRTSKNANSPVIEEQIPIYKQYIFRSDFSNEERFTFAFYAIAIDNEC 141
OY 143 OVAPGEALPTDHCNHLGFGYSCRAGYVLLHRNKRFCALSCGVFQSGELSPREY 202
Db 142 TDTF-DVP-CSHFCNNMTIGGFCSPPEYFLHDMKNGCVNCSGNVFTALIGELISSNYP 199
OY 203 RPYPKLSSCTYSISLEGFVILDFV-ESPDVETHPETLCPYDPLKIQOTREHGPFCGK 261
Db 200 NPYPENSRCFQILLEGFQVVVITQREDPDVBPADSGMGOQDILLPAKNRQGPFCGN 259
OY 262 TLPH--RIEKSNTVTTFTVDESQHTGKMTNSTAQCPCPYPMADPNGHVSGVQAKYI 319

```

```

Db 260 GFGPRLTIETHSNLTLDIVFQTDLEQKMKLRYHGDPICPREIT-ANSVMAPEKARY 318
OY 320 LKDSFISFCEGYELLGHLPLKSFYAVCOKDGSMDRPMACSIDCGPPDDLPGRVEX 379
Db 319 FKDVVKISCVDGEFAVGNVGSITFFYSTCOSNGOMSRLRCQPVDCIGLEPIONKVD- 377
OY 380 ITGPGVTYKAVIOYSCETFEYTK--VNDKYYVCEADGFWTSKGEKSLPVEPVCGLS 437
Db 378 -DPEMTLFCFSVIHYSCPEPYVMEHAEHGEYCAANGSVNDELIELPKPCVPVCVP 435
OY 438 AR--TTGGRTYGCKARPGDFPMQVYLIGGTAAGALLYNNVLTAAHAYEOKHDS-- 493
Db 436 TEPRRIQRIEFGEPRAIIOSEPMQV-FEPPRAGGALIGHWLTAAHVV-EGNSDPSMY 493
OY 494 --AIDIRMGTLKRLSPHYTQAMSEAFVHEGY-----TFDAGFNDIALIKLNKRVIN 545
Db 494 VGSTSVNENLANVQ-----KLITDRVLIHGMKRGDLSRTNFDNDIALVRLKDPVKMG 549
OY 546 SNITPICLPKREASFMRITDITAGSGWLTORGFLARNLMVYDIPVDHOKCTAAVEK 605
Db 550 PTVSPICLPQTSSEYEPSEGDGLISGMGRERNIVITQIRGAKLPVTSLEKROVKEEN 609
OY 606 PYPGGS--VTANMLCGLESGGCKDSCRGSGC--ALVFLDSETERFVGGIYSGSMNC 660
Db 610 PKRADPDYFTSNMTCAG--EKGVDSQDGSQGFALPVPNVDRPKRYVAGLVSMGK-- 664
OY 661 GEAGQYGVYTKVYIPIWI 679
Db 665 -KCGTYITIKVKNYKPMI 682

```

RESULT 5

C1HUS

Complement subcomponent C 1SBAR.GIF (EC 3.4.21.42) precursor [validated] - human

N:Alternate names: C1 esterase precursor

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Jun-2002

C:Accession: A40496; A27381; S00224; S26732; S05634; A05140; A25396; A38407; B37820

R:Kinoshita, H.; Hirose, S.; Salter, J.P.; Hagen, F.S.; Kurachi, K.

Proc. Natl. Acad. Sci. U.S.A. 85, 7307-7311, 1988

A:Title: Human genes for complement components C1r and C1s in a close tail-to-tail ar

A:Reference number: A40496; MUID:89017187; PMID:2459702

A:Molecule type: mRNA

A:Accession: A40496

A:Residues: 1-688 <RUS>

A:Cross-references: GB:J04080; NID:g179645; PIDN:AAA51852.1; PID:g179646

R:Tosi, M.; Duponchel, C.; Meo, T.; Jullier, C.

Biochemistry 26, 8516-8524, 1987

A:Title: Complete cDNA sequence of human complement C1s and close physical linkage of

A:Reference number: A27381; MUID:88163522; PMID:2831844

A:Accession: A27381

A:Molecule type: mRNA

A:Residues: 1-688 <ROS>

A:Cross-references: GB:M18767; NID:g179647; PIDN:AAA51853.1; PID:g179648

R:Mackinnon, C.M.; Carter, P.E.; Smyth, S.J.; Dunbar, B.; Fothergill, J.E.

Eur. J. Biochem. 169, 547-553, 1987

A:Title: Molecular cloning of cDNA for human complement component C1s. The complete a

A:Reference number: S00224; MUID:88082788; PMID:3500856

A:Accession: S00224

A:Molecule type: mRNA

A:Residues: 1-688 <MAC>

A:Cross-references: EMBL:X06596; NID:g29542; PIDN:CAA29817.1; PID:g763110

A:Accession: S26732

A:Molecule type: protein

A:Residues: 16-36;68-116,170-236,246-262,265-280,282-284,287-308,315-363,384-394,421-

R:Tosi, M.; Duponchel, C.; Meo, T.; Couture-Tosi, E.

J. Mol. Biol. 208, 709-714, 1989

A:Title: Complement genes C1r and C1s feature an intronless serine protease domain cl

A:Reference number: S05634; MUID:90040704; PMID:2553984

A:Accession: S05634

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 356-513, 'G', 514-688 <TO2>

R:Cartier, P.E Dunbar, B.; Forthergill, J.E.
 Biochem. J. 215, 565-571, 1983
 A>Title: The serine proteinase chain of human complement component C1s. Cyanogen bromide
 A:Reference number: A05140; MUID:84104122; PMID:6362651
 A:Accession: A05140
 A:Molecule type: Protein
 A:Residues: 438-483, 'X', 485-500:503-534:542-558:561-572, 'A', 574-601:617-623:626-644:647-
 Eur J. Biochem. 156, 49-57, 1986
 R:Spycher, S.E.; Nick, H.; Rickli, E.E.
 A>Title: Human complement component C1s. Partial sequence determination of the heavy chain
 A:Reference number: A25396; MUID:86164350; PMID:3007145
 A:Accession: A25396
 A:Molecule type: Protein
 A:Residues: 16-61:168-219:287-293, 'K', 295-334:384-445 <SP>
 R:Hess, D.; Schaller, J.; Rickli, E.E.
 Biochemistry 30, 2827-2833, 1991
 A>Title: Identification of the disulfide bonds of human complement C1s.
 A:Reference number: A38407; MUID:91175725; PMID:2007122
 A:Accession: A38407
 A:Molecule type: Protein
 A:Residues: 131-134, 'X', 136-146, 'X', 148-150:155, 'X', 157-162:166-170, 'X', 172-174, 'X', 176-
 187-402, 'X', 404-408:416-424, 'X', 426-431:547-556:592-597:617, 'X', 619-627, 'X', 629-633
 R:Thielens, N.M.; Aude, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.
 J. Biol. Chem. 265, 14469-14475, 1990
 A>Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated NH
 A:Reference number: A37820; MUID:90354439; PMID:2387866
 A:Accession: B37820
 A:Molecule type: Protein
 A:Residues: 16-25, 'X', 203-207 <TH>
 R:Thielens, N.M.; Van Dorsselaer, A.; Gagnon, J.; Arlaud, G.J.
 Biochemistry 29, 3570-3578, 1990
 A>Title: Chemical and functional characterization of a fragment of C1s containing the ep
 A:Reference number: A32672; MUID:90283368; PMID:2141278
 A:Contents: annotation: erythro-beta-hydroxyparagine site, content
 A:Note: about half of the A chains contain erythro-beta-hydroxyparagine
 C:Comment: This protein is a serine proteinase that combines with C1q and C1r to form C1
 S C2 and C4.
 C:Comment: C1s is a dimer of identical chains, each of which is activated by cleavage int
 C:Genetics:
 A:Gene: GDB:C1S
 A:Cross-references: GDB:119730; OMIM:120580
 A:Map position: 12p13-12p13
 A:Insertions: 291/1: 329/3: 356/1: 399/1: 424/1
 A:Note: the list of introns may be incomplete
 C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
 C:Keywords: acute phase; beta-hydroxyparagine; calcium binding; complement pathway; d
 C:15/Domains: signal sequence #status predicted <SIG>
 F:438-668/Product: complement subcomponent C1s chain B (light chain) #status experimental
 F:437-668/Product: complement subcomponent C1s chain A (heavy chain) #status experimental
 F:437/Product: complement subcomponent C1s chain A (heavy chain) #status experimental
 F:135-171/Domains: EGF homology <EGF>
 F:175-287/Domains: C1r/C1s repeat homology <C1R2>
 F:294-354/Domains: complement factor H repeat homology <FH1>
 F:359-421/Domains: complement factor H repeat homology <FH2>
 F:438-668/Product: complement subcomponent C1s chain B (light chain) #status experimental
 F:438-668/Product: complement subcomponent C1s chain A (heavy chain) #status experimental
 F:165-83, 135-147, 143-156, 158-171, 175-202, 234-251, 294-341, 321-354, 359-403, 366-421, 425-549,
 F:114/Modified site: erythro-beta-hydroxyparagine (Asn) (partial) #status experimental
 F:174/406/Binding site: carboxydrate (Asn) (covalent) #status experimental
 F:437-438/Cleavage site: Arg-116 (complement subcomponent C1r) #status experimental
 F:475,529,632/Active site: His, Asp, Ser #status predicted

[illegible]

A:Residues: 'S',27-39,'S',41,'T',43-54,'E',691-782,950-977 <TOR>
 C:Superfamily: coagulation factor C; C-type lectin homology; complement factor H repeat
 C:Keywords: alternative splicing; glycoprotein; hemolymph coagulation; hydrolase; serine
 F:1-25/Domain: signal sequence #status predicted <SI0>
 F:26-600/Product: coagulation factor C heavy chain #status experimental <RCH>
 F:126-195/Domain: complement factor H repeat homology <FH01>
 F:199-254/Domain: complement factor H repeat homology <FH02>
 F:260-321/Domain: complement factor H repeat homology <FH03>
 F:436-564/Domain: C-type lectin homology <LCH>
 F:576-634/Domain: complement factor H repeat homology <FH04>
 F:685-747/Domain: complement factor H repeat homology #status atypical <FH05>
 F:763-1019/Product: coagulation factor C light chain peptide A #status experimental <PPA>
 F:763-1019/Product: coagulation factor C light chain peptide B #status experimental <PPB>
 F:523,534,624,912/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:767/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:809,865,966/Active site: His, Asp, Ser #status predicted

Query Match 12.5%; Score 471.5; DB 2; Length 1019;
 Local Similarity 28.4%; Pred. No. 1.8e-23;
 Matches 152; Conservative 61; Mismatches 190; Indels 133; Gaps 27;

265 HRIET-----KNTV---TTF-VTDESDHTG-----WKIHTSTA 297
 Db 500 HRLDAEKFWEIMDRSNVYLDNLTFWAGSEPGNETNCYLDIRDQLQPYWK--KSCF 557

298 QP-----CPYMAPRGVHSPVQAKYILKDSFISFCETGELLQGHLL 341
 Db 558 QPSEFACMLDLSRNNKCKDDPGLNGHAT-LHGQSI--DGFAGSIRYSCVLYHLS 614

342 KSFYAVCQKDSMDRPMAC-SIVDC-GPPDDLPS-GRVEYITGPGVTT----- 387
 Db 615 GTEVTCCTGTGTSAPKPKRICIKVITCONP--VPSYGSVE-IKPRSTNSISRYGPELR 671

388 -----YAAVIO--YSCETFTMAVND 407
 Db 672 LPLRLPLARAKPRPRSSQSPSTVDLASKVYLRGCHYRGSRATYICERYELLGSG 731

408 GKVCYCEADGFWTSSKGEKSLPVEPCVGLSARTGRIYGGQAKKPPDFPMOVL----- 462
 Db 732 GRR-CDNGNMGSGRPAS-----CIPVCGRSDSPSPFIMGNSTNEICOWPQACISRWLA 785

463 ---LGTTAAGALLYDMVLTAAHAV---YEOKHNASALDIRMGTLK---LSPHYTQA 512
 Db 786 DHNMWFLQCGSLNLEKMIYTAACHVYSATVAETIDPSQFIYIGKYYRDSRDDYQV 845

513 MSEAVFHEGYTHDAGPDN-DIALIKLNKRVINSNTPICLPRK-EAESMRPDDIGTA 570
 Db 846 -REALETHVNPNDPGNLNFDIALIQLKTPVTLTRVQPICLPDTTTRHLEKGTAAV 904

571 SGNGLTORGFELARNLMLVDPIDVHOKCTAAYEKPRYPRGSVTANMLCAGLESQKDSR 630
 Db 905 TGMGLNNNTYSEMIQAVLPVVAASCTCEGKREADLP-LTVTNMCAAGYKKRYDACS 963

631 GDSGGLAVFL-DSETF-RMFVGVIGVSGS--MNCGEAGQGVYTKVINYPIENII 683
 Db 964 GDSGGLAVFADDSRTERRVLEGIVSGSPGCGKANQYGGFTKVVNVLFWIRQFI 1019

RESULT 9
 T30338
 Oviductin (EC 3.4.21.-) - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T30338; A40242
 R:Sunday, L.L., Wieduwilt, M.J.; Hedrick, J.L.
 Biol. Reprod. 60, 989-995, 1999
 A:Title: Oviductin, the Xenopus laevis oviductal protease that processes egg envelope g
 composed of two protease and several CUB domains.
 A:Reference number: 220830; M0ID:99184825; PMID:10084976
 A:Accession: T30338
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

A:Residues: 1-1004 <LIN>
 A:Cross-references: EMBL:U01291; NID:g1754713; PID:g1754714; PIDN:AA853972.1
 R:Hardy, D.M.; Hedrick, J.L.
 Biochemistry 31, 4466-4472, 1992
 A:Title: Oviductin. Purification and properties of the oviductal protease that proces
 A:Reference number: A40242; M0ID:92256375; PMID:1581303
 A:Accession: A40242
 A:Molecule type: protein
 A:Residues: 46-73 <HAR>
 C:Comment: This protease is found in oviductal secretory granules and is secreted to
 C:Superfamily: trypsin homology
 C:Keywords: extracellular protein; hydrolase; serine protease

Query Match 12.4%; Score 467; DB 2; Length 1004;
 Best Local Similarity 24.2%; Pred. No. 3.6e-23;
 Matches 181; Conservative 89; Mismatches 235; Indels 242; Gaps 33;

4 LTLGLLGSVAVPLGKWPVEF--GRLASPGF----- 35
 Db 252 LTSWGMCGCR-----SMKNVFLPHNRKSGPCIFTDIKLGWSSQLNTAVPNKNOE 304

36 -----PGE---YANDERRMTLAPPGYRLR-----LYFTHPDELSHLC 73
 Db 305 SCSMQDVLGKSGELLPLNPMVSTRMSGAPGFSLSLKTCTSCNLNFTHIDIESDFACN 364

74 YDFYKL-SSGAKVYLATLGGESTDTERAPGKDFYSLGSSLDITFRSDYSNEKPFTEEA 132
 Db 365 LDYLAITDTHRLGKFCGD-----IPRSLIISF-SSIKLWFSDFENR--TFVL 414

133 FYAEDIDECOVAPGEAPTCDDHCHNLGGRFYSCRAGVYLHRNKKRCSALCSGOVETOR 192
 Db 415 YLSA-----VEBNTVP--DSCG---GSP-----AVLEEE- 438

193 SGELSPETPRPKLSSCTYSISLEGFSYILDVFSFVETHPRLCPYDFLKIOTD- 251
 Db 439 -GEIQSMKYPENYLSNRCHMIIHGPGSYIKLOF-EDFALE--PSDCHSDSYLAVYODL 494

252 --REHGPRFCGKTLPHRIETKSNVTITFVTDSEGDHTGKIHYSIAOCPYMAPNG 309
 Db 495 AAEKIEFCGFSLPAYVYSTTAAVMIKFTDEDNOKGRATTT----- 539

310 HVSFVQAKYILKDSFISFCETGYELLQGHLLPLKSFYAVCQKDSMDRPMACSIYDCGP 369
 Db 540 FVSP-----NSLVEDS-----RQGNMP----- 556

370 DDLPSGRVEYITGPGVTTAAVIOYSCETFTYKAVNDKTVCEADGFWTSSKGEKSLPV 429
 Db 557 -----STKKKETTA 565

430 CEPVCGLSARTTGGRIYGC---QKAKPGDFPMOVL--LGTTAAGALLYDMNVLTAAH 483
 Db 566 QDSICGVY-QVPPIFYINSLAKVEAVPHSNPHITSLQYAGENHCDAIIAEMNILLTAS 624

484 AVEYEQK-HDASALDIRMGTLKRLSPHYTQAMSEAVFHEGYTHDAGFDNALIKLNKV 542
 Db 625 CVLNRKFNVDVLPDGIHDLR--PGHQKGVLQKIIIPHPSSQTN--DPIALVELDESL 682

543 VINSNTPICLPRKEASFPRTDIDGTASGNGLTQRFPLAR--NLAVDPIDVHOKCT 599
 Db 683 QFNSDIPICLPKCTSE--LAPASLCVYSGMSL--RGEAKKSTKLOQREVPIITDDACS 738

600 AAYEKPRYPRGSVTANMLCAGLESQ-GKDSCRGDSGALVFLDSETRMFWFGVIGVSGSM 658
 Db 739 AHYIQNP---GGITDRMLCAGIGGQNDSCSEDSGSPVLCLEKKGIYITFIASMG-V 794

659 NCGEAGQGVYTKVINYPIENIISD 685
 Db 795 NCKENSKPGIYTKVSPIDMIRQIMSD 821

RESULT 10
 J07731
 membrane-bound arginine-specific serine proteinase precursor - rat


```

Db 316 NPGIIRIFS--NQYATATFLIQ--SDESDYIGKVTYTAFTNSKELNNYKINCFNEDGFCFW 372
QY 137 -EDIDE-----CQVAREAPTCOHCHNHNLGFCYSCRAG-----170
Db 373 IQDLINDNEMERTGSGTPPTGTFPHTEGNE--SGFYISPTGPGGRERVELLTPLD 431
QY 171 -----VYLHNKRTCSALCS-----GOVFTORSGE- 195
Db 432 PTPQACLSFYVMYGEVNVYKLSINISSDQNMEXTIFQKEGNGONNNYGVYLTLEVEEF 491
QY 196 -----195
Db 492 KVSFYGKNOILSDIALDISLTFYICNVSVYRPTLVPTPPPLPTDCGPHDLMEPNT 551
QY 196 -LSSPEYPRPYPKLSCTYSISLEEGFSVILDFVESFDPVENHRETLCPYD-----FLK 247
Db 552 TETISINPNSTPNOAFCLWNLNAQKNIQHFQE--FDLENIADVETIRDEGDDSLFLA 610
QY 248 IQDTRREHGPCGKTLPHRIETKSNTVTITFEVDESQDHTGKIMHYTS-----TAQPCPY 302
Db 611 VYTG---PGPY-----NDVESTNRMVTLEITDNLAKQCFKANTTGYGLIPEQCK- 660
QY 303 PMAPNGHVSQAKYILKDSFISFCETGYELLQHLPLKSFYAVCQ-----KDSMDRP 357
Db 661 -----EDNFQ---CKDG-----ECIPLVN---LCDGFPHCKGSDSE-- 690
QY 358 MPACSIYDCGPPDLLPSGRVXYITGPGVTYTKAVIOYSCETFTMKVNDCKYVCEADGF 417
Db 691 -AHVRVLENGTSTD--SGLVQF-----RQISIMHVAQEN--WTTQISD--VCQLLGL 737
QY 418 WTSSKGEKSLPVE-----PYCGLSARTTG--443
Db 738 GT---GMSVYFTSTGGGPRVYVNLMTAPNGSILPLPSQOCLEDSLILQCNKSKGKLYT 794
QY 444 -----RIYGGOKAKPGDPWQOV--LILGGTTAGA--LYDNKVTLTAHAHYEQKHDA-- 493
Db 795 QEVSPIKIVGSGDSREGAMPVVALYFPDQOVCASLVSRLDVLVAACVYGRNNEPESKWK 854
QY 494 -ALDIRRGTLKRLSPHTQAMSEAVFIHEGYTHAGFDNDIALKLKNKVVINSITPIC 552
Db 855 AVGLGHAAS--NLTSFQIETRLIDQIVINPHY--NKRKNNDIAMMHLKKNVYTYTQIC 912
QY 553 LPRKEASFMTDDIGTASGNG--LTORGFLARNLMYVDIPVDHOKCTAAVEKPPYRGS 611
Db 913 LP--EENQVPPRGRIASIGAGALITYGSPADVLQEADVPLLSNEKC-----QQQMEBYN 965
QY 612 VTANMLCAGLESGKDCRQDGSALVFLDSETERFVGGIVSGSNMGAGAGYGYTK 671
Db 966 ITEMNVAGYEAGGVDCQSGGSGPL--MCOENNRMLLAGVTSFG--YQCALPNRPGVYAR 1022
QY 672 VINIPIENI 683
Db 1023 VPRTEMIOSTL 1034

```

RESULT 21

T31069

tolloid-BMP-1 like protein 1 - California sea hare

N:Alternate names: Probable metalloprotease TBL-1

C:Species: Aplysia californica (California sea hare)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 18-Aug-2000

C:Accession: T31069

R:Lit: O.R.; Hattar, S.; Endo, S.; MacPhee, K.; Zhang, H.; Cleary, L.J.; Byrne, J.H.; ES

A:Title: A developmental gene (Tolloid/BMP-1) is regulated in Aplysia Neurons by treatme

A:Reference number: 420965; PMID:98007484; PMID:8987797

A:Status: Preliminary;

A:Molecule type: translated from GB/EMBL/DBJ

A:Residues: 1-1070 <LID>

A:Cross-references: EMBL:U57369; NID:g1899041; PID:g1899042; PIDN:AAC47485.1

C:Superfamily: dorsal-ventral patterning protein tolloid; astacin homology; C1r/C1s repd

```

Query Match 10.8%; Score 406.5; DB 2; Length 1070;
Best Local Similarity 25.2%; Pred. No. 4e-19;
Matches 156; Conservative 69; Mismatches 236; Indels 137; Gaps 33;

```

QY 19 GPKWPEVFRSLSPGPFGEYANDQERRWTLTAPRGYRLRYTFHEDLSHCEDYEVK 78

Db 521 GGNMPDPE--GFLNSPAPVDEGSDKVCSEWITVEGQVALLFATFEEDPPCAYDYVE 579

QY 79 LSSG-----AVVLTLCQESTDPERARCKDTFSLSSLDITRSDYSNKKPTGFPAFY 134

Db 580 IRDDTKDSPVLTGYCG-----TRTPPNA--ISTSHLVTKVSDSMQK--GGFSASY 629

QY 135 AAEDICQVAPRGAEPYCDHCHNHNLGFCYSCAGYVLAHRNRTCSALCSGOVFTORS 194

Db 630 -LEEVDECE---GDRHCENVCVLTGSLYECTCKIGVLEHSDGKCKKACGGTL--DAPSG 664

QY 195 ELSSPEYPRPYPKLSCTYSISLEEGFSVILDFVESFDPVENHRETLCPYFLKIQT--D 251

Db 685 TISSPSEPDLYPPDKNCVMIISAPKGHITLVNFTH--MDLEMRDE--CELDVFRVTNVVGN 742

QY 252 REE--HGPFCKTLPHRIETKSNTVTITFEVDESQDHTGKIMHYTSQPCPYMAPNGH 310

Db 743 KERLOGQYCGFMAPPSTLSLSNELIEFRSDDTLQKTFGFSMDVYADVDEC---ASSNG- 797

QY 311 VSPYQAKYILKD--SFSICETGYELLQHLPLKSFYAVCOKDGSMDRPMACSIYDCG 367

Db 798 -----GCKHICENYVSGHSCGREGFILADE---KS-----CKEGGCHYEYTDTRKGVQSP 846

QY 368 P-PDDLPSGR--VEYITGPG-----VTTKYAVIOYSCF---EFFYTKMVD--GKY 410

Db 847 DYPSFYARRRDCENHFTTARGHVRLIFTDQVEPRHRTCYRDVHEAFDAGANIQAOLQKX 906

QY 411 VCEADGFWTSSKGEKSLPVC-----EKPAPIISSENTLITFLSPTSVQRKGFARHHDVCCSPTATISA-- 954

Db 907 -CGS-----EKPAPIISSENTLITFLSPTSVQRKGFARHHDVCCSPTATISA-- 954

QY 446 YGGOKAKPGDPWQOV--LILGGTTAGA--LYDNKVTLTAHAHYEQKHDA-- 493

Db 955 -----PKKIL--SHVLXGSRPYDNRNCSNITAP---BGCH---VELRFT 992

QY 501 TLKRLSPHTQAMSEAVFIHEGYTHAGFDNDIALKLKNKVVINSITPICLPRKEAES 560

Db 993 AFE--IEQOSRCLYDYAVVYDGP-----ENDLVL-----GKFCGNOVPSPIVSTSLV 1041

QY 561 FMRTDDIGTASGNGLQOR 578

Db 1042 RFRSDDTIKSGGFSATYR 1059

RESULT 22

KFHU

coagulation factor IXa (EC 3.4.21.22) precursor [validated] - human

N:Alternate names: antithromphillc factor B; Christmas factor

C:Species: Homo sapiens (man)

C:Date: 17-Dec-1982 #sequence_revision 30-Jun-1987 #text_change 15-Sep-2000

C:Accession: A00922; A37570; A30511; A32989; A22637; A21337; A37546; A30623; A60486;

R:Yoshitake, S.; Schach, B.G.; Foster, D.C.; Davie, E.W.; Kurachi, K.

A:Title: Nucleotide sequence of the gene for human factor IX (antithromphillc factor B

A:Reference number: A00922; PMID:8600558; PMID:2994716

A:Accession: A00922

A:Molecule type: DNA

A:Residues: 1-461 <YOS>

A:Cross-references: GB:K02402; NID:g182612; PIDN:AAB59620.1; PID:g182613

R:Anson, D.S.; Choo, K.H.; Rees, D.J.G.; Giannelis, F.; Gould, K.; Huddleston, J.A.;

EMD J. 3, 1053-1060, 1984

A:Title: The gene structure of human anti-haemophilic factor IX.

A:Reference number: A37570; PMID:84236100; PMID:6329734

A:Accession: A37570

A:Molecule type: DNA

A:Residues: 1-461 <ANS>

A:Cross-references: GB:K02048

R:Reitsma, P.H.; Bertina, R.M.; Ploos van Amstel, J.K.; Riemens, A.; Briet, E.
 Blood 72, 1074-1076, 1988
 A>Title: The putative factor IX gene promoter in hemophilia B Leyden.
 A:Reference number: A30511; PMID:88327116; PMID:3416069
 A:Accession: A30511
 A:Molecule type: DNA
 A:Residues: 8-24 <R&E>
 A:Cross-references: EMBL:X55008; NID:9311288; PIDN:CA838245.2; PID:94469253
 R:Koeberl, D.D.; Bottema, C.D.K.; Buerstedde, J.M.; Sommer, S.S.
 Am. J. Hum. Genet. 45, 448-457, 1989
 A>Title: Functionally important regions of the factor IX gene have a low rate of polymorphism.
 A:Reference number: A32889; PMID:89371752; PMID:2773937
 A:Accession: A32889
 A>Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 30-92 <KOE>
 R:McGraw, R.A.; Davis, L.M.; Noyes, C.M.; Lundblad, R.L.; Roberts, H.R.; Graham, J.B.; S
 Proc. Natl. Acad. Sci. U.S.A. 82, 2847-2851, 1985
 A>Title: Evidence for a prevalent dimorphism in the activation peptide of human coagulat
 A:Reference number: A22673; PMID:85190593; PMID:3857619
 A:Accession: A22673
 A:Molecule type: mRNA
 A:Residues: 1-193, 'T', 195-461 <MCG>
 A:Cross-references: GB:M11309; NID:9180552; PIDN:AA52023.1; PID:9180553
 A>Note: The authors translated the codon ACA for residue 29 as Tyr
 R:Jaye, M.; de la Salle, H.; Schamber, F.; Balland, A.; Kohl, V.; Findel, A.; Tolstosh
 Nucleic Acids Res. 11, 2325-2335, 1983
 A>Title: Isolation of a human anti-hemophilic factor IX cDNA clone using a unique 52-bp
 A:Reference number: A21337; PMID:83220788; PMID:6687940
 A:Accession: A21337
 A:Molecule type: mRNA
 A:Residues: 1-193, 'T', 195-461 <JAY>
 A:Cross-references: GB:00137; NID:9182610; PIDN:AA52763.1; PID:9182611
 R:Jagadeeswaran, P.; Lavell, D.E.; Kaul, R.; Mohandes, T.; Warren, S.T.
 Somat. Cell Mol. Genet. 10, 465-473, 1984
 A>Title: Isolation and characterization of human factor IX cDNA: identification of Tag I
 A:Reference number: A37546; PMID:84300526; PMID:6089357
 A:Accession: A37546
 A:Molecule type: protein
 A:Residues: 38-193, 'T', 195-326 <JAG>
 A:Cross-references: GB:M35672
 R:Kurachi, K.; Davie, E.W.
 Proc. Natl. Acad. Sci. U.S.A. 79, 6461-6464, 1982
 A>Title: Isolation and characterization of a cDNA coding for human factor IX.
 A:Reference number: A30623; PMID:83065193; PMID:6559130
 A:Accession: A30623
 A:Molecule type: mRNA
 A:Residues: 112, 'S', 14-73, 'P', 75-82, 'K', 84-203, 'P', 205-216, 'G', 218-298, 'A', 299-356, 'A',
 A:Cross-references: GB:00136; NID:9182608; PIDN:AA98726.1; PID:9182609
 R:Tharakan, J.; Strickland, D.; Burgess, W.; Drohan, W.N.; Clark, D.B.
 Vox Sang. 58, 21-29, 1990
 A>Title: Development of an immunofluorescent process for factor IX purification.
 A:Reference number: A60486; PMID:90194857; PMID:2316207
 A:Accession: A60486
 A:Molecule type: protein
 A:Residues: 47-52, 'XX', 55-60, 'X', 62, 'XX', 65 <THA>
 R:Kumlien, B.A.; Fujikawa, K.; Kistiel, W.
 Biochem. Biophys. Res. Commun. 115, 8-14, 1983
 A>Title: The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood co
 A:Reference number: A20274; PMID:83308813; PMID:6688526
 A:Accession: A20274
 A:Molecule type: protein
 A:Residues: 105-109, 'X', 111-115 <MCN>
 R:Balland, A.; Faure, T.; Carvallo, D.; Cordier, P.; Ulrich, P.; Fournet, B.; de la Salle
 Eur. J. Biochem. 172, 565-572, 1988
 A>Title: Characterisation of two differently processed forms of human recombinant factor
 A:Reference number: S02527; PMID:88166735; PMID:3280312
 A:Accession: S02527
 A:Molecule type: protein
 A:Residues: 29-63 <BAL>
 A>Note: processed forms expressed in recombinant system
 R:Ballat, S.; Petraud, F.; Dalemans, W.; Balland, A.; Dieterle, A.; Faure, T.; Meulien,

EMBO J. 9, 3295-3301, 1990
 A>Title: Characterization of recombinant human factor IX expressed in transgenic mice
 A:Reference number: S12058; PMID:91006024; PMID:2209546
 A:Accession: S12058
 A:Molecule type: protein
 A:Residues: 1-68 <JAL>
 A>Note: processed forms expressed in recombinant system
 R:Hanford, P.A.; Barton, M.; Mayhew, M.; Willis, A.; Beesley, T.; Brownlee, G.G.; Cam
 EMBO J. 9, 475-480, 1990
 A>Title: The first EGF-like domain from human factor IX contains a high-affinity calc
 A:Reference number: S12377; PMID:90151623; PMID:2406129
 A:Accession: S12377
 A:Molecule type: protein
 A:Residues: 92-130 <HAN>
 A>Note: NMR detection of calcium binding by domain expressed in recombinant system
 R:de la Salle, C.; Charmanier, J.L.; Baas, M.J.; Schwartz, A.; Miesel, M.L.; Grunbe
 Thromb. Haemost. 70, 370-371, 1993
 A>Title: A deletion located in the 3' non translated part of the factor IX gene respo
 A:Reference number: I59612; PMID:94054330; PMID:8236150
 A:Accession: I59612
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 444-461 <RES>
 A:Cross-references: GB:S66752; NID:9439773; PIDN:AAB28588.1; PID:9439774
 R:Stoffel, E.S.; Koeberl, D.D.; Sarkar, G.; Sommer, S.S.
 Science 239, 491-494, 1988
 A>Title: Genomic amplification with transcript sequencing.
 A:Reference number: I59529; PMID:88127096; PMID:3340835
 A:Accession: I59529
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 290-359 <RE2>
 A:Cross-references: GB:M19063; NID:9182622; PIDN:AA52456.1; PID:9182623
 R:Agarwala, K.L.; Kawabata, S.; Takeo, T.; Murata, H.; Shimomishi, Y.; Nishimura, H.;
 Biochemistry 33, 5167-5171, 1994
 A>Title: Activation peptide of human factor IX has oligosaccharides O-glycosidically
 A:Reference number: A54255; PMID:94227047; PMID:8172892
 A:Accession: A54255
 A:Molecule type: protein
 A:Residues: 'D', 204, 'X', 206-211; 212, 'D', 214, 'X', 216-221, 'D', <AC>
 A>Note: the residues designated 'X' were determined to be threonine bound to carbohyd
 R:Di Scipio, R.G.; Kurachi, K.; Davie, E.W.
 J. Clin. Invest. 61, 1528-1538, 1978
 A>Title: Activation of human factor IX (Christmas factor).
 A:Reference number: A18483; PMID:78194509; PMID:659613
 A:Contents: annotation; activation; active site; carbohydrate binding
 R:McGraw, R.A.; Davis, L.M.; Noyes, C.M.; Graham, J.B.; Roberts, H.R.; Stafford, D.W.
 Am. Soc. Hematol. Abstr. 64(Suppl.1), 262a, 1984
 A:Reference number: A37569
 A:Contents: annotation
 A>Note: 194-Thr was also found
 R:Morita, T.; Isase, B.S.; Esmon, C.T.; Johnson, A.E.
 J. Biol. Chem. 259, 5696-5704, 1984
 A>Title: Derivatives of blood coagulation factor IX contain a high affinity Ca2+-bind
 A:Reference number: A37543; PMID:84185715; PMID:6425296
 A:Contents: annotation; calcium binding
 R:Morita, T.; Isase, B.S.; Esmon, C.T.; Johnson, A.E.
 J. Biol. Chem. 260, 2583, 1985
 A:Reference number: A37544
 A:Contents: annotation; calcium binding; correction
 R:Bentley, A.K.; Rees, D.J.G.; Rizza, C.; Brownlee, G.G.
 Cell 45, 343-348, 1986
 A>Title: Defective propeptide processing of blood clotting factor IX caused by mutat
 A:Reference number: A37545; PMID:86189947; PMID:3009023
 A:Contents: annotation; signal sequence cleavage site
 R:Suehiro, K.; Kawabata, S.I.; Miyata, T.; Takeya, H.; Takamatsu, J.; Ogata, K.; Kami
 J. Biol. Chem. 264, 21257-21265, 1989
 A>Title: Blood clotting factor IX (F9) Nagoya: substitution of arginine 180 by trypto
 A:Reference number: A30622; PMID:90078229; PMID:2592373
 A:Contents: annotation; sequence of mutant B(M) Nagoya
 A>Note: carboxylation, glycosylation, and cleavage sites
 R:Barton, M.; Norman, D.G.; Harvey, T.S.; Hanford, P.A.; Mayhew, M.; Tse, A.G.D.; Brow
 submitted to the Brookhaven Protein Data Bank, November 1991

A:Reference number: A51252; PDB:1IXA
A:Contents: annotation, conformation by (1)H-NMR, residues 92-130
A:Note: recombinant form expressed in yeast
C:Comment: Factor IX is activated by factor XIa, which excises the activation peptide propeptide
C:Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K dependent gamma-carboxylation
C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with stromelysin, forms the active site
C:Genetics:
A:Gene: GDB:F9
A:Cross-references: GDB:119900; OMIM:306900
A:Map position: Xq27.1-xq27.2
A:Introns: 30/1; 84/2; 93/1; 131/1; 174/1; 241/3; 280/1
C:Function:
C:Description: catalyzes the proteolytic activation of coagulation factor X in the presence of factor V
A:Pathway: blood coagulation intrinsic pathway
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutamic acid
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-46/Domain: propeptide #status experimental <PP>
F:51-91/Domain: Gla domain homology <GLA>
F:128-191/Product: coagulation factor IXa light chain #status experimental <ALC>
F:128-191/Product: EGF homology <EGF>
F:128-191/Product: EGF homology <EGF>
F:128-226/Domain: activation peptide #status experimental <ACT>
F:227-461/Product: coagulation factor IXa heavy chain #status experimental <AHC>
F:227-461/Domain: trypsin homology <TRY>
F:53-54,61,63,66,67,72,73,76,79,82,86/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental <G>
F:59,69,97,108,102-117,119-128,134-145,141-155,157-170,178-335,252-268,382-396,407-435/Modified site: carboxylate (Ser) (covalent) #status experimental
F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F:191-192/Cleavage site: Arg-Ala (coagulation factor Xla) #status experimental
F:203,213/Binding site: carboxylate (Asn) (covalent) #status experimental
F:205,215/Binding site: carboxylate (Thr) (covalent) #status experimental
F:226-227/Cleavage site: Arg-Val (coagulation factor Xla) #status experimental

Query Match	10.8%	Score 405	DB 1	Length 461
Best Local Similarity	30.5%	Pred. No. 1.9e-19		
Matches 107	Conservative 50	Mismatches 92	Indels 102	Gaps 16

```

0Y 396 CEFETYYTKAKVADKVCACDGEFMTSSKGEKTLPCVCEPV---CG-TLSRFTG----- 442
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 141 CEG---FCRNSADNKKVYGCSTGTYLAEKOKS---CEPAVFPFGCRVSVSOTSKLTRAEV 195
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 443 -----GTYGOKAKRQDEPFOVLLIGTFA---AGA 471
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 196 FPDVDVYNSTAEETLLDNIITOSTGSEFNDFTRVVSGDEAKPOEFMOVYVLLGKADAFEGGS 255
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 472 LLYDNMVLTAAHAV-----YEOKHASALDIRMGLTKRLSPHYTG 511
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
0Y 256 IYNEMKIYTAACHVETGKIIIVVAGENHIEETHTEDKRN-----VNIIPHHN- 304
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 512 AMSEAVFHEGTYHDAGFDNDIALIKLNKKYVINSNTPICLPKKEASESMRTDIDIGTAS 571
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 305 -YNAAI---NKNKH-----DIALELEDEPLVMSYTPICIDAKETYNLFLEFGSOYVS 354
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 572 GNG-FTORGFLARNIMYDIPIDHOKCTAYAEKPPYPRGSVATNMICALESGGKOSCP 630
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 355 GNGRFFHNGRSALVLOYLRYPLVDRARCLRS-----TKFTIYNMKCAHFHGGSDSCQ 408
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 631 GDSGALVFLDSETE-RWYVCGIYSGMSGMCNGEGGYTKTYNTPWTE 680
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 409 GDSGSPHH---TEVEGTSFLLGIIISWE-ECAMGKKGAIYTKTSRYVYNIWK 455
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

RESULT 23

thrombin (EC 3.4.21.5) precursor - mouse
C.Species: Mus musculus (house mouse)
C.Date: 14-Dec-1990 #sequence_revision 14-Dec-1990 #text_change 03-May-2002
C.Accession: A35827; A42696; S12081
R.Degen, S.J.F.; Schaefer, L.A.; Jamison, C.S.; Grant, S.G.; Fitzgibbon, J.J.; Pal, J.A.
DNA Cell Biol. 9, 487-498, 1990
A.Title: Characterization of the cDNA coding for mouse prothrombin and localization of t
A.Reference number: A35827; MUID:91025551; PMID:222810

A:Accession: A35827
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-618 <DE>
A:Cross-references: NID:953813; PIDN:CA36548.1; PID:953814
A:Experimental source: strain C57BL/6
A>Note: the data were obtained from females resulting from the cross of *M. domesticus*
R.Banfield, D.R.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A:Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and
A:Reference number: A42696; MUID:92212913; PMID:1557383
A:Accession: A42696
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 384-618, 'E' <BAN>
A:Cross-references: GB:M81394
A:Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C:Keywords: blood coagulation; calcium binding; carboxyglutamic acid; glycoprotein; h
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-43/Domain: propeptide #status predicted <PRO>
F:28-88/Domain: Gla domain homology <GLA>
F:44-618/Product: prothrombin B #status predicted <MAT>
F:109-187/Domain: kringle homology <KR1>
F:215-293/Domain: kringle homology <KR2>
F:361-610/Domain: trypsin homology <TRY>
F:50-51, 58-60, 63-64, 69-70, 73-76/Modified site: gamma-carboxyglutamic acid (Glu) #stat
F:61-66, 91-104, 109-187, 130-170, 158-182, 215-293, 236-276, 264-288, 333-479, 388-404, 533-54
F:403, 459, 555/Active site: His, Asp, Ser #status predicted

Query Match	10.88;	Score 405;	DB 2;	Length 618;
Best Local Similarity	25.68;	Pred. No. 2.7e-19;		
Matches 163; Conservative	72;	Mismatches 232;	Indels 170;	Gaps 29

```

0Y 131 EAFYAEDIDECQYAPGAPRPPCDH-----HCNHNGRYCCRAGYVLAHRN 176
Db 70 EAFEELESPQDIDVFAWYATVCDVSRKPRETFMDCLEGRCAMDGVNVL----- 118
0Y 177 KRTCSALCSGOVFTQORS---ELSSPEYPRPYPKLSCTQYSLSLEBFSVILDFVESFDV 233
Db 119 -----GYVNYTHHGIOGQIMRSRPRHK-PELNSTTHGCA-----DKENFCR 159
0Y 234 ETHPEITLCPYDFLKIQIDREEH--GPFCCKTLPHRIETKSNTVYITFVTDESDGHGWKI 291
Db 160 NPDSSITGPMCVYTTDPYRRECSVPVCG-----OEGRTVYVMPRSGSGSKDN----- 207
0Y 292 HTYTAQACPRPMPAPRNHNHSPVQAKYILKQSFISFCETGELLQGLPLKSTFANG--- 348
Db 208 -----LSPPLG-----OCLTERG-RITQGNLAVTTGLSGPCLPW 239
0Y 349 -----QKDGSMDBRPACSIYD--CGPDDLPSCRVEYITG-PEVITYKAVIQYSC 397
Db 240 NSLPAKTLSTKQDDP---PEVKYLVENFCRNNDMBEGAMCYVAGQPDPEY-CULAN-CE 294
0Y 398 ETEYTMKVNDRKYVCE-----ADGFWTSKGERKSLPVCERVCGL-----SARTT--- 441
Db 295 EA-----VGEENYDVDESIAGRITDAEFHFFPNEKTFGLGADGCLRPLPEKSKLDYTE 349
0Y 442 -----GGRYGGOKRKBQDFPMOYVILIGT-----TAAGALLDMWVYLTAAHAV----- 485
Db 350 KELLDSDYIDGRIVGEMDAEKIAIOMVYALFPRKSOELLGCAISLIDRWYLTAAHCILYPP 409
0Y 486 YEOKHIDASALDIRMGTLKRLSPNHTQ-----AMSEAVFIHGXYHDGPFNDIALIKLN 540
Db 410 WDKNTTENDLVRIG--KHSTRIRERNVEKISMLEKTYIYHPRYRMRENDRDIALLLKX 467
0Y 541 KVINSNITPILCPKPE--AESFMRTDIDIGASGGLTQRFELAR-----NLMYVDIPT 592
Db 468 PVPFSDYTHPCLCPDKQCVTSLTLAGYKGRVATGNGNLRMTTININEIQPSVLOVWNLPI 527
0Y 593 VDHQKCTAAYEKPRPRCSVYANMLCAGL---ESGKDCSCRBGSGGALYVLDESETEKMFV 649
Db 528 VERPYCKAS-----TRITIDNNECGAGFYKVDNRKDDACDGGSGGFVAKSPFNKNWY 581

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 22:02:16 : Search time 82 Seconds
(without alignments)
1114.755 Million cell updates/sec

Title: US-09-874-198-2
Perfect score: 3758
Sequence: 1 MRLTLGLGCVATPLCP.....GYTFKVINYPWENIISDF 686

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 75 summaries

Database :

A.Geneseq-101002:*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3758	100.0	686	23	AAE14564 Human MASP-2 prote
2	3745	99.7	686	23	AAE14568 Human MASP-2 prote
3	3673	97.7	671	23	AAE14565 Human mature MASP-
4	1617	43.0	1034	22	ABG21134 Novel human diago
5	1608	42.8	728	22	AAAB47559 Protease PRS-1.
6	1600	42.6	728	22	AAAB85060 Human serine prote
7	1508	40.1	284	22	AAAB85075 Amino acid sequenc
8	1414	37.6	705	23	AAAG80757 Human C1r protein.
9	1412	37.6	705	22	AAAB50288 Complement compone
10	1364	36.8	707	23	AAAG80756 Murine C1r protein

ALIGNMENTS

RESULT 1	AAE14564	standard: Protein: 686 AA.	11	1238	32.9	760	21	AAAB3579	Human cancer assoc
			12	852.5	22.7	449	21	AAAB58274	Lung cancer associ
			13	681.5	18.1	351	21	AAAB58273	Lung cancer associ
			14	494.5	13.2	293	22	AAAB85077	Amino acid sequenc
			15	493.5	13.1	464	21	AAAB23624	Human secreted pro
			16	492	13.1	487	22	AAU086864	Human FcγR5b poly
			17	491.5	13.1	487	22	AAU086863	Human FcγR5a poly
			18	490	13.0	855	20	AAU06671	Tumour antigen der
			19	490	13.0	855	22	AAE06930	Human membrane-ty
			20	490	13.0	855	22	AAAB98500	Human TAD6-15. Ho
			21	488	13.0	762	21	AAAB19551	Human matrixpase (
			22	488	13.0	855	21	AAAY0284	Human peptidase, H
			23	488	13.0	855	21	AAAB19552	Human matrixpase.
			24	488	13.0	855	22	AAAB35465	Human membrane-ty
			25	486	12.9	851	22	AAAB11428	Human membrane-ty
			26	486	12.9	851	22	AAAB25628	Human protein sequ
			27	470.5	12.5	1019	18	AAAB43394	Singapore horsesho
			28	470.5	12.5	1019	20	AAAY05750	Horseshoe crab fac
			29	470.5	12.5	1019	20	AAAY4302	Horseshoe crab fac
			30	470.5	12.5	1019	21	AAAY42490	Recombinant N-term
			31	470.5	12.5	1019	22	AAAB60935	Horseshoe crab rec
			32	470.5	12.5	1083	18	AAAB43393	Singapore horsesho
			33	470.5	12.5	1083	20	AAAY05749	Horseshoe crab fac
			34	470.5	12.5	1083	20	AAAY94301	Horseshoe crab fac
			35	470.5	12.5	1083	21	AAAY42489	Horseshoe crab rec
			36	470.5	12.5	1083	22	AAAB60934	Horseshoe crab rec
			37	469.5	12.5	855	23	AAE23083	Epithin protein.
			38	455	12.1	296	22	AAAB85076	Amino acid sequenc
			39	452	12.0	932	22	AAAB21442	Novel human diago
			40	451	12.0	280	23	AAAB47910	MASP-like serine p
			41	450.5	12.0	1031	23	AAAB60993	Novel human protei
			42	447.5	11.9	296	22	AAAB85078	Amino acid sequenc
			43	446.5	11.9	269	22	AAAB84203	Amino acid sequenc
			44	445.5	11.9	902	22	AAAB98507	Murine epithin. M
			45	445.5	11.9	902	23	AAAB7549	Murine type II mem
			46	445.5	11.9	902	23	AAAB80517	Mouse epithillin-11
			47	444	11.8	802	20	AAAY41710	Human PRO618 prote
			48	444	11.8	802	21	AAAB44266	Human PRO618 (UN3
			49	444	11.8	802	21	AAAB24052	Human PRO618 prote
			50	444	11.8	802	23	AAAB82755	Amino acid sequenc
			51	443	11.8	658	22	AAE06934	Human membrane-ty
			52	443	11.8	802	22	AAE06933	Human membrane-ty
			53	443	11.8	1019	22	AAE06940	Human enterokinase
			54	427	11.4	719	21	AAAB09927	Murine TANGO 215 p
			55	426.5	11.3	720	22	AAAB85892	Mouse serine prote
			56	424.5	11.3	970	23	AAAB82743	Amino acid sequenc
			57	424.5	11.3	1464	22	AAAB71111	Drosophila melanog
			58	421	11.2	1379	22	AAAB58860	Drosophila melanog
			59	414.5	11.0	509	22	AAAB31373	Amino acid sequenc
			60	413	11.0	737	22	AAAB93670	Human protein sequ
			61	413	11.0	737	22	AAAB85891	Human serine prote
			62	413	11.0	762	22	AAAB85893	Human serine prote
			63	411.5	10.9	986	23	AAAB90755	Human Tumour Endot
			64	410.5	10.9	986	18	AAAB13670	C-proteinase encod
			65	410	10.9	415	20	AAAY03203	Bovine enterokinase
			66	408.5	10.9	798	15	AAAB57283	Amino acid sequenc
			67	407.5	10.8	462	12	AAAB10868	Recombinant human
			68	406.5	10.8	436	13	AAAB22513	Truncated precursor
			69	406	10.8	461	19	AAAB40284	Human Factor IX pr
			70	405.5	10.8	356	23	AAAB76529	Human LP polypepti
			71	405	10.8	454	16	AAAB67710	Human Factor-IX.
			72	405	10.8	456	5	AAAB40178	Part of the sequen
			73	405	10.8	461	6	AAAB50311	Sequence of human
			74	405	10.8	461	6	AAAB50302	Sequence of human
			75	405	10.8	461	11	AAAB05393	Mutant human Facto

XX AAEL1564;
AC 17-MAY-2002 (first entry)
DT Human MASP-2 protein.
DE Human MASP-2 protein.
XX Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL;
XX complement fixation; infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
OS Homo sapiens.
XX
XX Key
XX Location/Qualifiers
XX Peptide
XX 1..15
XX /label= Signal_peptide
XX 16..686
XX /label= Mature_MASP-2_protein
XX 16..56
XX /note= "MASP-2 20 kDa fragment"
XX 16..31
XX /label= C1r/C1s-like_domain
XX 16..296
XX /note= "This region is specifically claimed in claim 46"
XX 30..444
XX /note= "This region is specifically claimed in claim 41."
XX The region 30 to 296 is specifically claimed in claim 47"
XX 108..134
XX /note= "Fragment obtained by amino acid sequencing
XX of peptides"
XX 135..146
XX /label= EGF-like_domain
XX 138..296
XX /note= "This region is specifically claimed in claim 43"
XX 155
XX /note= "Encoded by CA"
XX 156
XX /note= "Encoded by C"
XX 183..199
XX /label= C1r/C1s-like_domain
XX 293..307
XX /label= CCP-1_domain
XX /note= "Complement control protein domain"
XX 363..376
XX /label= CCP-2_domain
XX 377..388
XX /note= "Fragment obtained by amino acid sequencing
XX of peptides"
XX 410..417
XX /note= "Fragment obtained by amino acid sequencing
XX of peptides"
XX 432..441
XX /note= "Linker"
XX 445..468
XX /label= Serine_protease_domain
XX 483
XX Active-site
XX 532
XX Active-site
XX 633
XX
XX WO200206460-A2.
XX
XX 24-JAN-2002.
XX
XX 13-JUL-2001; 2001WO-DK00499.
XX
XX 13-JUL-2000; 2000DK-0001089.
XX 01-JUN-2001; 2001DK-0000670.
XX
XX (JENS/) JENSENIUS J C.
XX (THIE/) THIEL S.
XX
XX Jensenius JC, Thiel S;

DR WPI: 2002-179791/23.
DR N-PSDB: AAD24224.
XX
XX Use of a polypeptide comprising amino acid sequence derived from
XX mannan-binding lectin associated serine protease-2 (MASP-2) for
XX producing pharmaceutical composition, to treat bacterial, fungal, viral
XX infections
XX
XX Claim 41; Fig 6; 76pp; English.
XX
XX The invention relates to use of a polypeptide derived from
XX mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
XX producing a pharmaceutical composition. MASP-2 is a complement-
XX fixing enzyme and involved in lectin pathway of complement activation.
XX The pharmaceutical composition comprising MASP-2 is useful for
XX treating infections caused by microbes such as fungus, yeast,
XX retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
XX bacteria which are resistant to at least one antibiotic medication or
XX multiresistant. The polynucleotide encoding MASP-2 is useful for
XX treating patients deficient in MASP-2. The invention also discloses
XX MASP-2 assays which are useful for determination of MASP-2 activity or
XX levels in patients suffering from e.g. infections, inflammatory disorders
XX and spontaneous recurrent abortion. The pharmaceutical composition
XX comprising MASP-2 inhibitor is useful for treating inflammatory
XX disorders. The present sequence is human MASP-2 protein.
XX Note: The present sequence is listed as being the same as
XX SEQ ID NO:2 shown in sequence listing of the specification (AAEL1568).
XX However the sequences differ at various locations.
XX
XX Sequence 686 AA:
XX
XX Query Match 100.0%; Score 3758; DB 23; Length 686;
XX Best local Similarity 100.0%; Pred. No. 1.5e-215;
XX Matches 686; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MRLTLGLGCVATPLGPKMEPEVGRGLASPGFGEYANDDERMTLTAPGRRLRY 60
XX |
XX 1 MRLTLGLGCVATPLGPKMEPEVGRGLASPGFGEYANDDERMTLTAPGRRLRY 60
XX
XX 61 FTHFDELSHCEYDFVKLSGAKVLATLGOESTDTERAPGKDFYSLGSSLDITFRSD 120
XX |
XX 61 FTHFDELSHCEYDFVKLSGAKVLATLGOESTDTERAPGKDFYSLGSSLDITFRSD 120
XX
XX 121 YSNEKPTFGFARYAADDIDECVARGEARPTCDHCHNHLGCGYCCRAGYVLRKKKRC 180
XX |
XX 121 YSNEKPTFGFARYAADDIDECVARGEARPTCDHCHNHLGCGYCCRAGYVLRKKKRC 180
XX
XX 121 YSNEKPTFGFARYAADDIDECVARGEARPTCDHCHNHLGCGYCCRAGYVLRKKKRC 180
XX |
XX 181 SALCSGOVFTQKSGELSSPEYPRPKLSCTYSISLEEGFSYILDFVESFDVETPETL 240
XX |
XX 181 SALCSGOVFTQKSGELSSPEYPRPKLSCTYSISLEEGFSYILDFVESFDVETPETL 240
XX
XX 241 CPYDFLKIQTDRREHGFPGCKTLPHRIETKSNVTVTTFVDESDDHTGKIHVYTAQPC 300
XX |
XX 241 CPYDFLKIQTDRREHGFPGCKTLPHRIETKSNVTVTTFVDESDDHTGKIHVYTAQPC 300
XX
XX 301 PYMAPPNHVSFVQAKYILKDSFSPCEYGLLOGHLPLKFTAVCQKDSMDRPMRA 360
XX |
XX 301 PYMAPPNHVSFVQAKYILKDSFSPCEYGLLOGHLPLKFTAVCQKDSMDRPMRA 360
XX
XX 301 PYMAPPNHVSFVQAKYILKDSFSPCEYGLLOGHLPLKFTAVCQKDSMDRPMRA 360
XX |
XX 361 CSTVDCGPPDDLPSSGREVETYGVTYKAVIOYSCSEFTFYTKKVDGKYVCADSFMTS 420
XX |
XX 361 CSTVDCGPPDDLPSSGREVETYGVTYKAVIOYSCSEFTFYTKKVDGKYVCADSFMTS 420
XX
XX 421 SKGEKSLPVCPEYCGLSARTTGRIRYGGOKAKGDPFMOVLLIGTTAAGALLYDMWVLT 480
XX |
XX 421 SKGEKSLPVCPEYCGLSARTTGRIRYGGOKAKGDPFMOVLLIGTTAAGALLYDMWVLT 480
XX
XX 481 AAHAVYEQKHDSALDRIKGTTLKRLSPHTQAMSEAVFIHEGTHRAGDNDIALKLNN 540
XX |
XX 481 AAHAVYEQKHDSALDRIKGTTLKRLSPHTQAMSEAVFIHEGTHRAGDNDIALKLNN 540
XX
XX 541 KYVINSNTPICLPKREAESFMRTDDIGTASGGLORGLANLWYVDPIDYDHOKCTA 600
XX |
XX 541 KYVINSNTPICLPKREAESFMRTDDIGTASGGLORGLANLWYVDPIDYDHOKCTA 600
XX

Db 541 KVINSNITPICLPKRAESFMRTDIDGTASGWLQRGFLARNLMYVDIPVDHOKCTA 600
QY 601 AYKPPYPRGSVTANMLCAGLESCKGDCRGDSGALVFLDSETERMFVGGIYSWGSUNC 660
Db 601 AYKPPYPRGSVTANMLCAGLESCKGDCRGDSGALVFLDSETERMFVGGIYSWGSUNC 660
QY 661 GEAGQYGVYTKVINYIPWENIISDF 686
Db 661 GEAGQYGVYTKVINYIPWENIISDF 686
RESULT 2
AAEL14568
ID AAEL14568 standard; protein; 686 AA.
XX
AC AAEL14568;
XX
DM 17-MAY-2002 (first entry)
Human MASP-2 protein, alternative version.
Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
complement fixation; infection; microbe; retrovirus; HIV; abortion;
human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..15
FT Protein /label= Signal_peptide
FT /label= 16..686
FT MISC-difference 155 /label= Mature_MASP-2_protein
FT MISC-difference 156 /note= "Encoded by CA"
FT MISC-difference 156 /note= "Encoded by C"
FT MISC-difference 298 /note= "Encoded by CAG"
FT MISC-difference 299 /note= "Encoded by CAG"
XX
PN WO200206460-A2.
XX
PD 24-JAN-2002.
XX
PF 13-JUL-2001; 2001WO-DK0499.
XX
PI 13-JUL-2000; 2000DK-0001089.
XX
PA 01-JUN-2001; 2001DK-0000870.
XX
PA (JENS/) JENSENIUS J C.
XX
PA (THIE/) THIEL S.
XX
PI Jensenius JC, Thiel S;
XX
DR WPI; 2002-179791/23.
XX
DR N-PSDB: AAD24224.
XX
PT Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections
XX
XX
PS Claim 41: Page 71-73; 76pp; English.
XX
CC The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or

CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is useful for treating inflammatory
CC Note: The present sequence is stated as being the same as
CC SEQ ID NO:2 shown in figure 6 of the specification (AAEL14564).
CC However the sequences differ at various locations.
XX
SQ Sequence 686 AA:
Query Match 99.7%; Score 3745; DB 23; Length 686;
Best Local Similarity 99.7%; Pred. No. 8.9e-215;
Matches 684; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MRLTLLGLLGSVAPPLGPKWPEVGRLASPGFPEYANDQERRMTLTAPGRLRLY 60
Db 1 MRLTLLGLLGSVAPPLGPKWPEVGRLASPGFPEYANDQERRMTLTAPGRLRLY 60
QY 61 FTTHDELHLCEYDFYKLSGAKVLTLCQESTDTERAPGKDTFVSLGSSLDITFRSD 120
Db 61 FTTHDELHLCEYDFYKLSGAKVLTLCQESTDTERAPGKDTFVSLGSSLDITFRSD 120
QY 121 YSNKPFTEAFYAARDIDECQVAPGEAPPCDHCHNHGGFCSGRAGVILHRNRTC 180
Db 121 YSNKPFTEAFYAARDIDECQVAPGEAPPCDHCHNHGGFCSGRAGVILHRNRTC 180
QY 181 SALCSGVFTQSGELSSPEYPRPYPLSSCTYSISLEGFSVILDEVESFDVETHPETL 240
Db 181 SALCSGVFTQSGELSSPEYPRPYPLSSCTYSISLEGFSVILDEVESFDVETHPETL 240
QY 241 CPYDFELIQDREHGFPCGKTLPHRIETKSNVTYTFVDESQDHTGKIHVSTAHAC 300
Db 241 CPYDFELIQDREHGFPCGKTLPHRIETKSNVTYTFVDESQDHTGKIHVSTAHAC 300
QY 301 PYMAPPNGHVSPYQAKYIILKDSFISFCETGYELLOGLHLPLKSTTAVCOKDSMDRMPA 360
Db 301 PYMAPPNGHVSPYQAKYIILKDSFISFCETGYELLOGLHLPLKSTTAVCOKDSMDRMPA 360
QY 361 CSIYDCGPPDDLPSGRVEYITGPGVTTYKAVIOYSCBETFYMKVNDGKYVCEADGFWTS 420
Db 361 CSIYDCGPPDDLPSGRVEYITGPGVTTYKAVIOYSCBETFYMKVNDGKYVCEADGFWTS 420
QY 421 SKGKSLPVCCEPVCGLSARTTGGRIYGGOKAKPDQFPMQVLLIGCTTAAGALLYDNVLT 480
Db 421 SKGKSLPVCCEPVCGLSARTTGGRIYGGOKAKPDQFPMQVLLIGCTTAAGALLYDNVLT 480
QY 481 AAHAVYEQKHDAALDIRMGTLKRLSPHYQAMSEAVFIHGGYTHDGFNDIALILNN 540
Db 481 AAHAVYEQKHDAALDIRMGTLKRLSPHYQAMSEAVFIHGGYTHDGFNDIALILNN 540
QY 541 KVINSNITPICLPKRAESFMRTDIDGTASGWLQRGFLARNLMYVDIPVDHOKCTA 600
Db 541 KVINSNITPICLPKRAESFMRTDIDGTASGWLQRGFLARNLMYVDIPVDHOKCTA 600
QY 601 AYKPPYPRGSVTANMLCAGLESCKGDCRGDSGALVFLDSETERMFVGGIYSWGSUNC 660
Db 601 AYKPPYPRGSVTANMLCAGLESCKGDCRGDSGALVFLDSETERMFVGGIYSWGSUNC 660
QY 661 GEAGQYGVYTKVINYIPWENIISDF 686
Db 661 GEAGQYGVYTKVINYIPWENIISDF 686
RESULT 3
AAEL14565
ID AAEL14565 standard; peptide; 671 AA.
XX
AC AAEL14565;
XX
DM 17-MAY-2002 (first entry)

XX		Human mature MASP-2 protein.
DE		
XX		
KW		Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KV		Complement fixation; infection; microbe; retrovirus; HIV; abortion;
KX		human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
OS		
XX	Homo sapiens.	
FH		
FT	Key	Location/Qualifiers
ET	Region	15..671
FT		/note="This region is specifically claimed in claim 45"
XX		
PN	WO200206460-AZ.	
PD		
PF	24-JAN-2002.	
	13-JUL-2001; 2001WO-DK00499.	
	3-JUL-2000; 2000DK-0001089.	
	1-JUN-2001; 2001DK-0000870.	
PA	(JENS/) JENSENIUS J C.	
PA	(THIE/) THIEL S.	
XX	Jensenius JC, Thiel S;	
PI		
DR	WPI; 2002-179791/23.	
XX		
PT		
PT	Use of a polypeptide comprising amino acid sequence derived from	
PT	mannan-binding lectin associated serine protease-2 (MASP-2) for	
PT	producing pharmaceutical composition, to treat bacterial, fungal, viral	
PT	infections -	
XX		
PS	Claim 45; Page 73-76; 76pp; English.	
XX		
CC	The invention relates to use of a polypeptide derived from	
CC	mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for	
CC	producing a pharmaceutical composition. MASP-2 is a complement-	
CC	-fixing enzyme and involved in lectin pathway of complement activation.	
CC	The pharmaceutical composition comprising MASP-2 is useful for	
CC	treating infections caused by microbes such as fungus, yeast,	
CC	retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic	
CC	bacteria which are resistant to at least one antibiotic medicament or	
CC	multiresistant. The polynucleotide encoding MASP-2 is useful for	
CC	treating patients deficient in MASP-2. The invention also discloses	
CC	MASP-2 assays which are useful for determination of MASP-2 activity or	
CC	levels in patients suffering from e.g. infections, inflammatory disorders	
CC	and spontaneous recurrent abortion. The pharmaceutical composition	
CC	comprising MASP-2 inhibitor is useful for treating inflammatory	
CC	disorders. The present sequence is human mature MASP-2 protein.	
XX		
SQ	Sequence 671 AA:	
	Query Match 97.7%; Score 3673; DB 23; Length 671;	
	Best Local Similarity 99.7%; Pred. No. 1.7e-210;	
	Matches 669; Conservative 0; Mismatches 2; Indels 0; Gaps 0	
OY	16 TPGLGWPEPVGRLASPGFGEYANDQERRTLTAPGYRLLYFTTFADLELSHCXD 75	
Dd	1 TPLGWKEPPVGRILASPGFPGEYANDQRRRWLTLPAGYRLRYTPFDLELSHCXD 60	
OY	76 FKLSSGAKVALATTCGGESTDTERARPKRTFFSLGSLIDITFRSDVSNNKPFTGFAPAFA 135	
Dd	61 FKLLSSGAVALATTCGGESTDTERAPGKRTFFSLGSLIDITFRSDVSNNKPFTGFAPAFA 120	
OY	136 AEDIDECQVAPEAETCDHHCHNHLAGFCSCRAGVYLHRNKRRTCASCALSGGVTRSGE 195	
Dd	121 AEDIDECQVAPEARTCDHHCHNHLAGFCSCRAGVYLHRNKRRTCASCALSGGVTRSGE 180	
OY	196 LSPPEYPRPYPKLSSCTYSISLEEGFSVLDFVESFDEVETHRPETLCVDYFLAIQTDDREH 255	

Db	181	LSSEVPPPPYKRLSSCTYSISLEEGFVILDFEVSFVETHPETLCYDELKIQTDREH	240
Qy	256	GPRCKGLTPHHLEIKNSMNVTTPTPTDSSGHTGKIHVSTAOCPPVPMAPRPGHVS	315
Db	241	GPFCGLTPHHLEIKNSMNVTTPTPTDSSGHTGKIHVSTAACIPVPMAPRPGHVS	300
Qy	316	AKYLLKDSFSTFECETGYELLQGHLPKSFYAVCCQKDGSMWRPMPACSTVDCGPPDDL	379
Db	301	AKYLLKDSFSTFECETGYELLQGHLPKSFYAVCCQKDGSMWRPMPACSTVDCGPPDDL	366
Qy	376	RVEYITGPGVTTTAAVIOYSCSEETFTMKVNDKRYVCEADGFWTSSKGEKSLPYCE	435
Db	361	RVEYITGPGVTTTAAVIOYSCSEETFTMKVNDKRYVCEADGFWTSSKGEKSLPYCE	420
Qy	436	LSARTGGRIRYGGGKAKPGDPPMOVLILGGTTAGALLYDMMVLTAAHAYEOKH	495
Db	421	LSARTGGRIRYGGGKAKPGDPPMOVLILGGTTAGALLYDMMVLTAAHAYEOKH	480
Qy	496	DIRMGTLKRLSPHYTQAMSEAVFTHEGYTHADGFDNDIALIKLNKVVINSNTPTCL	555
Db	481	DIRMGTLKRLSPHYTQAMSEAVFTHEGYTHADGFDNDIALIKLNKVVINSNTPTCL	540
Qy	556	KEASFMTDIDIGTASGMLTQRGFLARNLMYVDIPYDHQKCAAAEKRPYPGSGVTAN	615
Db	541	KEASFMTDIDIGTASGMLTQRGFLARNLMYVDIPYDHQKCAAAEKRPYPGSGVTAN	600
Qy	616	MLCAGLESGCKDSCRGDSGALVFLDSETERMFVGGLVSMGSMNCGEAGQGYVTKV	679
Db	601	MLCAGLESGCKDSCRGDSGALVFLDSETERMFVGGLVSMGSMNCGEAGQGYVTKV	666
Qy	676	IPWIENTIISDF 686	
Db	661	IPWIENTIISDF 671	
RESULT 4			
ABG21134			
ID	ABG21134	standard; Protein: 1034 AA.	
XX	XX	ABG21134:	
AC	XX	18-FEB-2002 (first entry)	
XX	XX		
DE	XX	Novel human diagnostic protein #21125.	
XX	XX		
KW	XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	XX	food supplement; medical imaging; diagnostic; genetic disorder.	
XX	OS	Homo sapiens.	
XX	XX	W0200175067-A2.	
PN	XX		
PD	XX	11-OCT-2001.	
XX	XX		
XX	XX	30-MAR-2001; 2001WO-US08631.	
XX	XX		
PR	XX	31-MAR-2000; 2000US-0540217.	
PR	XX	23-AUG-2000; 2000US-0649167.	
PA	XX	(HYSE-) HYSEQ INC.	
PI	XX		
DR	XX	Drmamac RT, Liu C, Tang YT;	
DR	XX	WPI: 2001-639362/73.	
XX	XX	N-PSDB: AAS85321.	
XX	XX		
PT	XX	New isolated polynucleotide and encoded polypeptides, useful in	
PT	XX	diagnostics, forensics, gene mapping, identification of mutations	
PT	XX	responsible for genetic disorders or other traits and to assess	
XX	XX	biodiversity	
XX	XX		
XX	XX	Claim 20; SEQ ID NO 51493; 103bp; English.	

PD 27-SEP-2001.
 XX
 PF 16-MAR-2001; 2001MO-US08441.
 XX
 PR 17-MAR-2000; 2000US-190708P.
 PR 30-MAR-2000; 2000US-193182P.
 PR 14-APR-2000; 2000US-197086P.
 PR 20-APR-2000; 2000US-199022P.
 PR 28-APR-2000; 2000US-200227P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Lu DAM, Policky JL, Delegeane AM, Tribouley CM, Khan FA;
 PI Au-Young J, Bandman O, Lai P, Borowsky ML, Gandhi AR, Hillman JL;
 PI Tang YT, Burford N, Baughn MR, Nguyen DB, Yao MG, Walla NK, He A;
 PI Hafalla A, Lu Y, Paterson C;
 DR WPI; 2001-611509/70.
 DR N-PSDB; AAR43512.
 CC New polypeptide for treating gastrointestinal, cardiovascular and
 CC autoimmune disorders, comprises novel human proteases (PRTs) and
 CC polynucleotides -
 XX
 PS Claim 1; Page 109-11; 129pp; English.
 XX
 CC The sequences given in AAB47559-69 show novel human proteases PRTs-1
 CC to -11. These proteases can be administered in pharmaceutical compounds
 CC for the treatment of diseases or conditions associated with decreased
 CC expression of functional protease. Disorders which can be treated
 CC include gastrointestinal (peptic esophagitis, indigestion and
 CC gastritis), cardiovascular (hypertensive heart disease, myocardial
 CC infarction), autoimmune/inflammatory (rheumatoid arthritis), cell
 CC proliferative (arteriosclerosis, cancers), epithelial (eczema),
 CC neurological (Huntington's disease), and reproductive (endometriosis)
 CC disorders. This protein shows homology with the human precursor of
 CC p100 serine protease of Ka-reactive factor.
 CC
 XX Sequence 728 AA:
 SO
 Query Match 42.8%; Score 1608; DB 22; Length 728;
 Best Local Similarity 43.0%; Pred. No. 9.8e-88;
 Matches 307; Conservative 124; Mismatches 245; Indels 38; Gaps 11;
 QY 1 MRLTLGLCGVATPLGPKWP-EPVGRGLASPGFGEYAMQDERMTLTAPPGYRL 59
 1 MRMLLYALCFSLSKASATVELNMNFGQIQSPGPDVPSDSEVYWNITVDPGFRIKL 60
 60 YFTHFDLHSLCEYDVVKLSGAKVLAATCGQESTDTERAPGKDTFYSLGSSLDITFRS 119
 61 YFMHFNLESSYLCEYDVVKETEDQVLAFCGRETDTEDQPGQEVVLSGSEFMSITFRS 120
 QY 120 DYSNEKFTGFEAFYAEDIDECOVANGEPATCDHCHNLGFGYCSRAGYVLAHRKRT 179
 121 DFNSEKFTGFDAAHYMAVDYDECKERDELDCHDCHNTIGGYGCRGVLHTDNRK 180
 QY 180 CSAKSGQVTFORSGLSSPEYRPYKLSCTYSLISLEGFSVILDFVSFVETHPET 239
 181 CREVCSNLFORTGYTSPDPNPYKSECIYLTIELGSPVNVNLFQEDIPDIEDHPEV 240
 QY 240 LCYDPLKIQTDREHRRPFCGKTLPHRIETKSNVTITFTYDSDGHTGKIHSTQAQ 299
 241 PCPYDYIKIKVGRVLGPPCGEKAPETISTOSHVLILFHSNDSNGENRGLSYRAGNE 300
 QY 300 CFPVMAAPNCHVASPYOAKYLLKDSFSFCFTGYELLQGHLPKSFATVACQKDGSDWRPMP 359
 301 CPELOPPVHOKIEPSQAKYFFKQVLSCTGFKVLKLDNEMPTFQTECLKDGWSMKIP 360
 QY 360 ACSIVDCGPDDLPSSGVEYITGPVTTYKAVIYQSCETFFYTM-KVNDGKYCEADGFW 418
 361 TCAIVDCRAGCELEHGLITFTSTNNLTYSKELKYSQGEYKMLNNNTIYTCSAGQVW 420
 QY 419 TSSKGEKSLFVCEPVCGLSARTTGG--RIYGGQAKAPGDFPMOVLILGGTTA----- 468

DB 421 MNKVLGRSLPTCLPEGQPSRSLSLVKRIIGGRNAPGLFPMQALIVEDITSRVNDKW 480
 QY 469 --GALLIYDNWVLTAAHAYVEQKHDSALDI-RMGTLKRLSPHYTAQMSA-----VF 518
 DB 481 FGSGALLSASWILTAALVLSQRRDTPVIVSKREHVTYVIGLHDYDKSCAUVSSAARV 540
 QY 519 IHGYTHDAGFNDIALIKLNKRVINSNTIPCLPRKEASEFMRDIDGTASGWLTOR 578
 DB 541 LHPDF-NIQVYNNIDIALYQDEVPPLGPHMPVCLPRLEEG-PATHMLGLVAGWGISNP 598
 QY 579 GF-----LARNLWYDPIYDHOKTAAYEKPPYPGSVYANMLCAGLESCK 626
 DB 599 NTVDETISSGTRTSLDVLQYKPLPVPAHECKTSYESRS-GNYSVENMFCAGYEGCK 657
 QY 627 DSCRGDSGALVFLDSETERFVGGTVSGMS-MNCGAGQGYTKVINYIPAI 679
 DB 658 DTCIGDSGAFVIFPDLISQHWVQGLVSWGPECCSKQYGVYTVSNYGVW 711
 RESULT 6
 AAB85060
 ID AAB85060 standard; Protein; 728 AA.
 AC AAB85060;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Human serine protease MASP-3 polypeptide.
 XX
 KW Mannan-binding lectin associated serine protease-3; MASP-3; MBL; human;
 KW mannan-binding lectin; complement; infection; cancer; cytostatic;
 KW immunomodulator; neuroprotective; anticonvulsive; antirheumatic;
 KW antirheumatic; antianemic; antiinflammatory; antiporiatic; antidiabetic;
 KW antitherosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN W0200140451-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 30-NOV-2000; 2000MO-DK00659.
 XX
 PR 02-DEC-1999; 99DK-0001721.
 PR 21-JUL-2000; 2000DK-0001126.
 XX
 PA (JENS/) JENSENIUS J C.
 PA (THIE/) THIEL S.
 XX
 PI Jensenius JC, Thiel S;
 DR WPI; 2001-374820/39.
 DR N-PSDB; AAF83895.
 XX
 PT Novel pure mannan-binding lectin associated serine protease
 PT polypeptides and polynucleotides encoding the protein, used for
 PT treating reoxygented ischemic tissues, mannon-binding lectin
 PT deficiency, multiple sclerosis
 XX
 PS Claim 15; Page 88-96; 99pp; English.
 XX
 CC The invention relates to a mannan-binding lectin (MBL) associated serine
 CC protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on
 CC the complement activation particularly when bound to MBL/MASP-2 complexes
 CC and directly activates complement system through binding to MBL. The
 CC MASP-3 polypeptides, polynucleotides and modulators are useful for
 CC preparing a pharmaceutical composition for treating aberrant MASP-3
 CC activity such as infections, cancer, MBL-deficiency, disorders of the
 CC immune system and reproductive system. The MASP-3 polypeptides is also
 CC used for treating diseases associated with human immunodeficiency virus,
 CC multiple sclerosis, myasthenia gravis, epilepsy, rheumatoid arthritis,
 CC vasculitis, autoimmune hemolytic anemia, Crohn's disease, asthma,

CC diabetes, psoriasis, multiple myeloma, atherosclerosis etc. It is also
 CC useful for inhibiting activation of C4 complement by inhibiting the MBL
 CC pathway, inhibiting MASP-2 activity, inhibiting or treating an
 CC inflammatory condition related to complement activation through MBL/MASP
 CC complexes. MASP-3 is also useful for treating an inflammatory condition
 CC resulting from an autoimmune condition after acute myocardial infarction
 CC or brain ischemia. It is also useful for treating an individual suffering
 CC from a disorder resulting from an imbalanced cytokine network. The
 CC present sequence represents a human MASP-3 polypeptide.

XX
 SQ Sequence 728 AA;

Query Match 42.6%; Score 1600; DB 22; Length 728;
 Best Local Similarity 42.7%; Pred. No. 2.9e-87;

Matches 305; Conservative 125; Mismatches 246; Indels 38; Gaps 11;

QY 1 MRLTLGLGCGVATPLGPKWP-EPVFGRLASPGFPEYANDQERRWLTAPGGRRL 59
 1 MRLTLGLGCGVATPLGPKWP-EPVFGRLASPGFPEYANDQERRWLTAPGGRRL 60
 60 YFTHFDELSHLCGEYDFVKLSGAKVLAATLGGESTDTERAPKDFYSLGSSLDITFRS 119
 61 YFTHFDELSHLCGEYDFVKLSGAKVLAATLGGESTDTERAPKDFYSLGSSLDITFRS 120
 QY 120 DYSNEKPTGFEAFYAEDIDECOVAPGEAPTCDDHCHNHLGGFYSCRAGYVLRHNKRT 179
 121 DFNENERTGFDHAYMAVDDECKEREDELSCDHCNHNIGGYCSCRGYILHNDKRT 180
 QY 180 GMLGCGVFTQSGELSPFPRPKLSCTYSISLEGEVYIIDPVESEFVETHPET 229
 181 CRVECDNLTQRTGVTGTSDFPNPKSECLYTIELGEGFVNMIOFEDIFDIODHPEV 240
 QY 240 LCPYDFLKIOTDEEHPGCKTLPHRIETKSNVTITFTDDESGDTHKIHYSNAOP 299
 241 PCGYDIKIKVGRKVLGPGCEKAPERTISQSHSVLLPHSDSAERKRLRYRAAGNE 300
 QY 300 CRYPMAPNGHVSPOAKYILKDSFISFETGYELLQHLPLKSFYAVCCOKDSMDRPM 359
 301 CPGLQPPVHGKIEPSQAKYEFKDOVLVSCDTGYKVLKDNEMDTFQECCLKDTWSNKP 360
 QY 360 ACSIVCGPPDDLPSGRVEITGPGVTTKAVIYQVCEEFYTM-KVNDCKYCEADGFW 418
 361 TCYIVCRAPGELEHGLITFTSRNNLTYSKELTKSCQEPYVMNNNGIYTCSAQGVW 420
 QY 419 TSSKGEKSLPCEPVCGLSARTTG---RIYGQAKKPGDFPMQVILIGGTA----- 468
 421 MNKVLGRSLPTCLPECGGSRSLPSLVKRTIGGRNAEPGLFPMQALIVEDTISRVRNDK 480
 469 --AGALLYNVLTAAHAAYEQKHDAALDI-RMGTLKRLSPHYTQAMSEA-----VF 518
 481 FGSGALLASWILTAAHVLRSGRRDTTVPVSKENHTVYLGLDHVRDKSGAVNNSAARV 540
 QY 519 IHGCTHDAQFNDIALIKLNKVVINSNTPTCLPRKEAESFMRTDDIGTASGGLTQR 578
 541 LHPDF-NIQNYNHDIALVDQEPVPLGPHMPCVCLPRLEPEG-PADHMLGLVAGWGSINP 598
 QY 579 GF-----LARNLMYVDIPVDHOKCTAAEKPYPGSGTANMLGSLGSGCK 626
 599 NVTVDELIISGRTTSLDVIQYKLPVPAHECTSYESRS-GNYSTENMEFCAGYEKGK 657
 QY 627 DSCRGDSGALVFLDSETERFVGVIGVSWGS-MNCGEAGQYGVYTVINITYPI 679
 658 DTCIGDSGAFVIFDDLSQRWVQGLVSWGGPECGSKQYGYTVYSNVDMV 711

RESULT 7
 AAB83075
 ID AAB85075 standard; peptide; 284 AA.
 AC AAB85075;
 XX
 DT 06-AUG-2001 (first entry)

DE Amino acid sequence of MASP-2 polypeptide.

XX Mammalian-binding lectin associated serine protease-3; MASP-3; MBL;
 KW mannin-binding lectin; complement; infection; cancer; cytostatic;
 KW immunomodulator; neuroprotective; anticonvulsive; antineumatic;
 KW antidiabetic; antineurotic; antineurotic; antineurotic; human;
 KW antidiabetic; antineurotic; antineurotic; antineurotic; human;
 XX

OS Homo sapiens.

PN WO200140451-A2.

PD 07-JUN-2001.

PF 30-NOV-2000; 2000WO-DK00659.

PR 02-DEC-1999; 99DK-0001721.

PR 21-JUL-2000; 2000DK-0001126.

PA (JENS/) JENSENIUS J C.

PI (THIE/) THIEL S.

PI Jensenius JC, Thiel S;

DR WPI; 2001-374820/39.

PT Novel pure mannin-binding lectin associated serine protease

PT polypeptides and polynucleotides encoding the protein, used for

PT treating reoxygated ischemic tissues, mannin-binding lectin

PT deficiency, multiple sclerosis

PS Disclosure: Fig 6; 99pp: English.

CC The invention relates to a mannin-binding lectin (MBL) associated serine
 CC protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on
 CC the complement activation particularly when bound to MBL/MASP-2 complexes
 CC and directly activates complement system through binding to MBL. The
 CC MASP-3 polypeptides, polynucleotides and modulators are useful for
 CC preparing a pharmaceutical composition for treating aberrant MASP-3
 CC activity such as infections, cancer, MBL-deficiency, disorders of the
 CC immune system and reproductive system. The MASP-3 polypeptide is also
 CC used for treating diseases associated with human immunodeficiency virus,
 CC multiple sclerosis, myasthenia gravis, epilepsy, rheumatoid arthritis,
 CC vasculitis, autoimmune hemolytic anemia, Crohn's disease, asthma,
 CC diabetes, psoriasis, multiple myeloma, atherosclerosis etc. It is also
 CC useful for inhibiting activation of C4 complement by inhibiting the MBL
 CC pathway, inhibiting MASP-2 activity, inhibiting or treating an
 CC inflammatory condition related to complement activation through MBL/MASP
 CC complexes. MASP-3 is also useful for treating an inflammatory condition
 CC resulting from an autoimmune condition after acute myocardial infarction
 CC or brain ischemia. It is also useful for treating an individual suffering
 CC from a disorder resulting from an imbalanced cytokine network. The
 CC present sequence represents a MASP-2 polypeptide, used for comparison
 CC studies with MASP-3.

SQ Sequence 284 AA;

Query Match 40.1%; Score 1508; DB 22; Length 284;
 Best Local Similarity 98.6%; Pred. No. 3.3e-82;

Matches 282; Conservative 0; Mismatches 0; Indels 4; Gaps 2;

QY 403 MKYNDCKYVCEADGFWTSSGKSLPCEPVCGLSARTTGRTYGGOKAPGDFPMQVLT 462
 1 MKYNDCKYVCEADGFWTSSGKSLPCEPVCGLSARTTGRTYGGOKAPGDFPMQVLT 58
 QY 463 LGGTTAAGAL-LYDNMVLTAHAAYEQKHDAALDI-RMGTLKRLSPHYTQAMSEAVFT 520
 59 LGGTTAAGALLIYDNMVLTAHAAYEQKHDAALDI-RMGTLKRLSPHYTQAMSEAVFT 118
 QY 521 EGYTHDAGFNDIALIKLNKVVINSNTPTCLPRKEAESFMRTDDIGTASGGLTQRGF 580
 119 EGYTHDAGFNDIALIKLNKVVINSNTPTCLPRKEAESFMRTDDIGTASGGLTQRGF 178

PT gene -
XX Claim 23: Page 107-109; 140pp; English.
XX
XX The invention relates to methods for diagnosing and prognosing ovarian
CC tumors in an individual via the detection and measurement of the
CC expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,
CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
CC ABA83181 and ABA83183). The methods of the invention are useful for
CC detecting an ovarian tumour in a patient, for identifying an individual
CC at increased risk for developing ovarian cancer, in prognostic tests for
CC assessing the relative severity of ovarian cancer, in tests for
CC monitoring a patient in remission from ovarian cancer and in tests for
CC monitoring disease status in a patient being treated for ovarian cancer.
CC The methods can additionally be used to identify a particular tumour as
CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from
CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
CC mucinous cystadenoma, borderline mucinous tumour, mucinous
CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
CC tumour. The ovarian tumour marker genes of the invention were identified
CC using SAGE (serial analysis of gene expression) and were found to be
CC overexpressed in a broad variety of ovarian epithelial tumour cells
CC relative to normal ovarian epithelial cells. The marker genes are
CC implicated in immune response pathways, in the regulation of cell
CC proliferation and in protein folding, and many of these are membrane-
CC localised or secreted. In addition to their use as diagnostic and
CC prognostic markers, the ovarian tumour marker genes or their encoded
CC proteins may be used as therapeutic targets for the treatment and
CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent
CC proteins encoded by ovarian tumour marker genes of the invention.
XX
SQ Sequence 705 AA:

Query Match 37.6%; Score 1412; DB 22: Length 705;
Best Local Similarity 41.7%; Pred. No. 4, 2e-76;
Matches 301; Conservative 108; Mismatches 248; Indels 64; Gaps 22;

OY 1 MRLTLTLL--GLTC--GSVATPLGPKMPEVFGRLASPGFGEYANDQERMTLTAPPGY 55
DB 1 MMLTLLVLLALPFRAGST--PI-----PQKLREVTSPLEPKRYPNNEFTTYITPTGY 54
OY 56 RLRLFTTHFDELSHLCEYDFVKLSSGAVLATLGCQESTDTERAPKDTFVLSGLSDI 115
DB 55 RVLVRFQDFLEPSECGCFDYVVISADKKSILGRFCQLGSLNPGKKEFMGQKMLL 114
OY 116 TFRSDVSNKPP-----FTGFEAFYAADIDEC--QVAPGE--APTCDDHCHHNLGCFYC 165
DB 115 TFRSDVSNKPP-----FTGFEAFYAADIDEC--QVAPGE--APTCDDHCHHNLGCFYC 165
OY 166 SCRAGYVLLHRNKRKTCALSCGVFTORSGFLSPERPRPKLSCTSYSLSEEGFSVIL 225
DB 175 SCRAGYVLLHRNKRKTCALSCGVFTORSGFLSPERPRPKLSCTSYSLSEEGFSVIL 225
OY 226 DEYVESDVETHPETLCPYDFLKIQTDRREHGPFQCTLPRIETKSNVTYITFVTDSEGD 285
DB 235 KFLPEPDIDIDHOQVHCYPDQLQIYANGKNNGEPCGKQRPDLDTSSNAVDLFFETDESGD 284
OY 286 HTGNKLNHTSTQAPCPYMAPRN-GHVSVPQAKYIILKDSSTICEFYELQCHLPLKSF 344
DB 295 SKRMKRYTTEIKCPQPTLDEFTTIQNLQDPQYQFDFYIATCKQYQILBENQVLIHSP 354
OY 345 TAVCQKDSMDRPMRACSTVDCGPPDDLPSGREYITFGVTVYKAVIOVSCSEFTYTKM 404
DB 355 TAVCQKDSMDRPMRACSTVDCGPPDDLPSGREYITFGVTVYKAVIOVSCSEFTYTKM 404
OY 405 V-----NDGKYVCEADGF--TSSGKESLAPCEPVCG--ISARTTGGRIYGGQAKRP 454
DB 415 TRAGSRESEQGYVYTAQGIIMKNEQGEK-IPRCLPYCGKPVNVPQROQLIGQGAKKMG 473
OY 455 DPEWQVLIIGTTAAGALLYDNVNLTAHAHAYVQKHA---SALDLIRMG-----TLKRLS 506
DB 474 NPEWQVLIIGTTAAGALLYDNVNLTAHAHAYVQKHA---SALDLIRMG-----TLKRLS 506

OY 507 PHYTOAMSEAVFIEGTYHDA--GFDNDIALIKLNKVNINSITPTCLPRKEAESFMR 564
DB 534 NHPTR-----RVSHPPDYRQDESYNFEDDIALLENSVTLGPMLPLCLP--DNDFTYDL 587
OY 565 DDIGTASGWLTRGFLARNLMYDIPYIDHQCTAYEKPPYPRGS-----VTANMLCA 619
DB 588 GLMGVYSGFCVMEER-IAHDLRFVRLPVPANPQC-----ENMLRGNRMDSVQNNFCA 640
OY 620 GLESGKDSRGSGCALVLDSETERMFVGGIVSGSMNCGEAGQGYTKYINIPWT 679
DB 641 GHPSLQDQACQDSGCVFAVRDPTDRWATGIVSMG-IGCSRG--YGFYTKLVNVDWI 697
OY 680 E 680
DB 698 K 698

RESULT 10
AAG80756
ID AAG80756 standard; protein; 707 AA.
XX AAG80756;
AC AAG80756;
XX 15-APR-2002 (First entry)
XX Murine C1r protein.
XX Murine C1r protein.
XX Murine; C1r; tissue rejection; skin allograft.
XX Mus sp.
XX KR2001077614-A.
XX 20-AUG-2001.
XX 03-FEB-2000; 2000KR-0005516.
XX 03-FEB-2000; 2000KR-0005516.
XX 03-FEB-2000; 2000KR-0005516.
XX (BIOC-) BIO CLUE & SOLUTION CO LTD.
XX (KIMT/) KIM T Y.
XX Byun SJ, Hwang HY, Kim GE, Kim TY, Park YI, Ryu JU;
XX WPI; 2002-111648/15.
XX Mouse gene relating to rejection against skin allograft -
XX Disclosure: Page 9-10; 11pp; Korean.
XX This invention describes novel murine and human proteins, C1r, associated
XX CC with tissue rejection against skin allograft. This sequence represents
XX CC the murine C1r protein described in the invention.
XX
SQ Sequence 707 AA:

Query Match 36.8%; Score 1384; DB 23: Length 707;
Best Local Similarity 39.2%; Pred. No. 2e-74;
Matches 282; Conservative 120; Mismatches 260; Indels 58; Gaps 16;

OY 3 LTLTLGLGCVATPLGPKMPEVFGRLASPGFGEYANDQERMTLTAPPGYRLRLYFT 62
DB 7 LTLTLGLGCVATPLGPKMPEVFGRLASPGFGEYANDQERMTLTAPPGYRLRLYFT 62
OY 63 HFDLELSHLCEYDFVKLSSGAKVLAATLGCQESTDTERAPKDTFVLSGLSDITFRSDYS 122
DB 61 QDFVSESGCFDYVVISADKKSILGRFCQLGSLNPGKKEFMGQKMLLTFTPTDPS 120
OY 123 NEKP-----FTGFEAFYAADIDECOVAP-----GEAPTCDDHCHHNLGCFYCSCAGY 172
DB 121 NEKNGTIFMKGFLAYQAVDIDECASQPSNVEGLOPRCQCHLCHNYVGVCFCSCHPGYE 180

Db 592 PEGRTFNDNDIALVRLKDPYKMGPTVSPICLPETSSDYINLMDGLLISGWGTEKRDRA 651
QY 583 RNLMYVDIPDVHQK-TAAVEKPPYGRGS--VTANMLCAGLESBGKSGDSCRGSGALVF 639
Db 652 VRLKARLPVAPLRKCKEYKVEKPTADAEVYTPPMICAGGK-GMDSCKGSGGAFAY 710
QY 640 LD-SETERMFVGIVSGWSNMGCEAGQGYVYTKVINYIPWI 679
Db 711 QDPNDKTKFYAAGLVSMGP----QCGTGLYTRVKKNYVDWI 747

RESULT 12

AAB58274
ID AAB58274 standard; Protein: 449 AA.

AC AAB58274;
XX
XX

DT 14-MAR-2001 (first entry)

Lung cancer associated polypeptide sequence SEQ ID 612.

Human: lung cancer associated protein; neuroprotective; cytostatic;
cardioactive; immunomodulatory; muscular active; vulnery;
gastrointestinal; nephrotoxic; antiinfective; gynecological;
antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
proliferative disorder; wound healing; infectious disease.

OS Homo sapiens.

PN WO20005180-A2.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05918.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Ruben SM;

XX WPI; 2000-587514/55.

DR N-PSDB: AAF18150.

PT Lung cancer associated gene sequences, referred to as lung cancer
antigens, useful for treatment, prevention, and diagnosis of disorders
such as lung cancer -

Claim 11: Page 1104-1106; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
associated proteins represented in AAB58106 - AAB58548. Lung cancer
associated proteins and polynucleotide sequences, their agonists, and
antagonists may have neuroprotective; cytostatic; cardioactive;
immunomodulatory; muscular active general; vulnery; gastrointestinal
general; nephrotoxic; antiinfective; gynecological; or antibacterial
activity. The invention also includes antibodies specific for the
protein or polynucleotide sequences. The lung cancer associated
polynucleotide sequences may be used for detection of lung cancer,
chromosome identification, as chromosome markers, and for numerous other
diagnostic or research purposes. The proteins may be used to treat
disorders such as neural, immune, muscular, reproductive,
gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
disorders. The proteins may also be used in the treatment of wounds and
infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
peptide AAB58549 are used in the course of the invention for the
identification and characterisation of the polynucleotide and protein
sequences.

XX Sequence 449 AA;

Query Match

22.7%; Score 852.5; DB 21; Length 449;

Best Local Similarity 41.2%; Pred. No. 4, 8e-43;
Matches 186; Conservative 63; Mismatches 159; Indels 43; Gaps 15;

QY 256 GPRCGKTLPHRITKSNVTITVTDESBDHTGMKLNHYSTAPRCYRMAPRN-GIVSPV 314
Db 9 GEFCKQKRPDLDITSSNADLLEFTEBSGDSRGKRLYTEIILKCPQKTLDEFTIIQNL 68
QY 315 QAKYIILKDSFISICETGYELLGHLPLKSFPAVCOKGSDMRPMPCSIYDCGPPDLP 374
Db 69 QPOYQERDFTIATCKGQYOLIEGNOYLHSTTAVCDDGTHNRAMPKIKIDCCQPRLPN 128
QY 375 GRVEYITGPGVTYKAVIQYSCETFEYTMKV-----NDGKYVCEADGFW-TSSKGEKS 426
Db 129 GDFRYTTMGVNTYKARIQYCHEPYKMQTRAGSRESEGGVYTCYAOGITMKNKQGEK- 187
QY 427 LPYCEPCC--LSARTTGATYGGOKAKPCDFFPMOYLIGGTTAAGALLYDNNVLTAAHA 484
Db 188 IPRCLPYCGKPVNPVDQRITIGQKAKMGFPVOYFTNIGRGGALLLDRLMILTAAHT 247
QY 485 VYEQKHDA--SALDIRMG-----TLKRLSPHYTOAMSEAVFIHEGYTHDA--GFNDIA 534
Db 248 LYPKEHQAOSNADLVFLGHTNVEELMKLNHPIRRS-----VHPDYRODESTNFBEDIA 303
QY 535 LIRLNKRVINSITPICLPKKEAESFMRTDIGTASGWLGTGRFLARNLMYVDIPTVD 594
Db 304 LLELENSVTLGPNLIPICLP--DNDTFYDLGLMGVSGFGEVMEK-IAHDLRFVRLPVAN 360
QY 595 HQCTAAVEKPPYGRGS--VTANMLCAGLESBGKSGDSCRGSGALVLDSETEMFV 649
Db 361 PQAC-----ENMLRGKRNMDVFSQNNFCAGHPSLQDACGDSGCVFAVRDPTDRWA 414
QY 650 GGIVSMGSMNMGCEAGQGYVYTKVINYIPWI 680
Db 415 TGIVSMG-IGCSRG--YGFYTKVINYVDWIK 442

RESULT 13

AAB58273
ID AAB58273 standard; Protein: 351 AA.

AC AAB58273;
XX
XX

DT 14-MAR-2001 (first entry)

Lung cancer associated polypeptide sequence SEQ ID 611.

Human: lung cancer associated protein; neuroprotective; cytostatic;
cardioactive; immunomodulatory; muscular active; vulnery;
gastrointestinal; nephrotoxic; antiinfective; gynecological;
antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
proliferative disorder; wound healing; infectious disease.

OS Homo sapiens.

PN WO20005180-A2.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05918.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Ruben SM;

XX WPI; 2000-587514/55.

DR N-PSDB: AAF18149.

PT Lung cancer associated gene sequences, referred to as lung cancer
antigens, useful for treatment, prevention, and diagnosis of disorders
such as lung cancer -

XX Claim 11; Page 1103-1104; 1425pp; English.
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC antigonists may have neuroprotective; cytoskeletal; cardiovascular;
CC immunomodulatory; muscular active general; vulnarary; gastrointestinal
CC general; nephrotropic; antineoplastic; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.

Sequence 351 AA;

Query Match 18.1%; Score 681.5; DB 21; Length 351;
Best Local Similarity 43.0%; Pred. No. 5,4e-33;
Matches 138; Conservative 50; Mismatches 112; Indels 21; Gaps 7;

QY 1 MRLTL--GLLC--GSVATPLGPKMPVGRRLASPGFPGYANDQERMTLTAPRGY 55
DB 2 MWLLYLTPALFCRAGGS1-PI---POKLFGEVNSPLFPKPYPNNEFTTITVPTGTG 55
QY 56 RLRLYTHHDLTSLHCEDFVNLSSGAKLALTCQESTDTRRAGKDTFTYLSGLDI 115
DB 56 RKLVLVQGDLEBSECFDYKISADKSLGFCGOLSPGPNPGKKFEMSGCKMML 115
QY 116 TRSDYSNEKP----FTGEAFYAEDIDEC--QVAPGE--APTCDHHCHNLGGFYC 165
DB 116 TPTDTSNENGTIMFYKFLAYVQAVDLDECASRSKSGEDHPQOCMLCHNYVGGYFC 175
QY 166 SCRAQVILHRNKRKTSALCSGOVETORSGELSPREPRPKISCTGYSLSFEGSVIL 225
DB 176 SRRPGELEDBDRHSCOECSSELYTEASGYISLEYPRSTPPDLRCNYSIRVERGLTLHL 235
QY 226 DVESEDFVETHPETLCPYFLKIQTDREHGPFCKTLPHRIETKSNFTYTFVTDESGD 285
DB 236 KFLPEPDIDHQGVHCPYQQLIYANGKNIGECGKQRPDLDTSSNAVLDLFTDESGLD 295
286 HTGWMKIHYSIAOPCPYPMAP 306
296 SRGWMKLRYYTTEXHQVPPADP 316

RESULT 14

ID AAB85077 standard; peptide; 293 AA.

XX AAB85077;

DT 06-AUG-2001 (first entry)

DE Amino acid sequence of C1r polypeptide.

XX Mannan-binding lectin associated serine protease-3; MASP-3; MBL;
KW mannann-binding lectin; complement; infection; cancer; cytoskeletal;
KW immunomodulator; neuroprotective; anticonvulsive; antineumatic;
KW antidiabetic; antianemic; antiinflammatory; antipsoriatic; human;
KW antidiabetic; antithrombotic; C1r.

OS Homo sapiens.

XX WO200140451-A2.

PD 07-JUN-2001.
XX 30-NOV-2000; 2000WO-DK00659.
PF 02-DEC-1999; 99DK-0001721.
PR 21-JUL-2000; 2000DK-0001126.
PA (JENS/) JENSENIUS J C.
XX (THIE/) THIEL S.
PI Jensenius JC, Thiel S;
DR WPI: 2001-374820/39.
XX Novel pure mannan-binding lectin associated serine protease
PT polypeptides and polynucleotides encoding the protein, used for
PT treating reoxygated ischemic tissues, mannan-binding lectin
PT deficiency, multiple sclerosis
XX Disclosure; Fig 6; 99pp; English.

CC The invention relates to a mannan-binding lectin (MBL) associated serine
CC protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on
CC the complement activation particularly when bound to MBL/MASP-2 complexes
CC and directly activates complement system through binding to MBL. The
CC MASP-3 polypeptides, polynucleotides and modulators are useful for
CC preparing a pharmaceutical composition for treating aberrant MASP-3
CC activity such as infections, cancer, MBL-deficiency, disorders of the
CC immune system and reproductive system. The MASP-3 polypeptide is also
CC used for treating diseases associated with human immunodeficiency virus,
CC multiple sclerosis, myasthenia gravis, epilepsy, rheumatoid arthritis,
CC vasculitis, autoimmune hemolytic anemia, Crohn's disease, asthma,
CC diabetes, psoriasis, multiple myeloma, atherosclerosis etc. It is also
CC useful for inhibiting activation of C4 complement by inhibiting the MBL
CC pathway, inhibiting MASP-2 activity, inhibiting or treating an
CC inflammatory condition related to complement activation through MBL/MASP
CC complexes. MASP-3 is also useful for treating an inflammatory condition
CC resulting from an autoimmune condition after acute myocardial infarction
CC or brain ischemia. It is also useful for treating an individual suffering
CC from a disorder resulting from an imbalanced cytokine network. The
CC present sequence represents a C1r polypeptide, used for comparison
CC studies with MASP-3.

SQ Sequence 293 AA;

Query Match 13.2%; Score 494.5; DB 22; Length 293;
Best Local Similarity 39.5%; Pred. No. 5.9e-22;
Matches 115; Conservative 44; Mismatches 97; Indels 35; Gaps 13;

QY 408 GRVCEADGFW-TSSKGEKSLPYCEPVCG--LSARTTGGRIYGGOKAKRPDPVOYLIG 464
DB 13 GYVYTAAGIWKNEQGER-IPRCLPVCCKPVNPBQRIITGGOKAKKNGFPWQFTNI 71
QY 465 GTTAGALLYDNMWTALAAVYEQKDA--SALDIRG-----TLKRLSPHYTQMSA 516
DB 72 HGRGGALLGDRWILTAATLYPKHEHAQSNASLDVFLHTNVEELMKLGNPIR-----R 127
QY 517 VFTHGTYTHDA--GPDNDALILKLNKVVYNSINPILCPKEAFSEFMTDDIGTASGNG 574
DB 128 VSYHPYRQDESYNEFGDIALLENSYTLGPRLDPLCP--DNDFTYDLGLMGVSGFG 185
QY 630 RGDSGALVFLDSETERMFVGGIVSGWSMNCBAGQGYVTVYNTPIPIE 680
DB 239 QGDSGCVFAVRDPNDRWVATGIVSWG-IGCSRG--YGFYTVLVNLVDMWK 286

RESULT 15

ID AAB23624 standard; Protein; 464 AA.

```

XX  AAB23624;
AC
XX
XX  12-JAN-2001 (first entry)
DE
XX  Human secreted protein SEQ ID NO: 48.
XX
XX  Human; secreted protein; cytokine; cell proliferation;
XX  nutritional supplement; immune modulation; autoimmune disorder;
XX  haematopoiesis regulation; tissue growth; haemostasis; inflammation.
OS
XX  Homo sapiens.
FH  Key
FH  Peptide
FT  2..14
FT  /label= signal_peptide
FT  15..464
FT  /label= mature_protein
XX
XX  0200049134-A1.
XX  24-AUG-2000.
XX
XX  18-FEB-2000; 2000WO-US04340.
XX
XX  19-FEB-1999; 99US-0120680.
XX  23-APR-1999; 99US-0298733.
XX  17-AUG-1999; 99US-0149639.
XX  23-SEP-1999; 99US-0156866.
XX  01-OCT-1999; 99US-0157247.
XX  29-NOV-1999; 99US-0167822.
XX  29-NOV-1999; 99US-0167823.
XX  15-FEB-2000; 2000US-0298733.
XX
XX  (ALPH-) ALPHAGENE INC.
XX
XX  Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiak P;
XX
XX  WPI: 2000-549267/50.
XX  N-PSDB: AAA93124.
XX
XX  New secreted proteins and polynucleotides encoding them, which are
XX  derived from Homo sapiens, useful for therapy, diagnosis, and research,
XX  as well as nutritional sources or supplements.
XX
XX  Claim 57; Page 275-276; 309pp; English.
XX
XX  The present sequence is the sequence of a human secreted protein. Its
XX  DNA was isolated from an adult muscle cDNA library. The proteins
XX  and coding sequences of the invention can be used in the isolation of
XX  similar genes and proteins, in the elucidation of their function in vivo,
XX  and to treat a number of conditions. It is possible that they may have
XX  uses as nutritional supplements, as cytokine or cell proliferation
XX  factors, in immune modulation, where they may be used to treat immune
XX  and autoimmune diseases, as haematopoiesis regulators (treating myeloid or
XX  lymphoid cell deficiencies), in the promotion of tissue growth, they may
XX  have chemokine or chemotactic activity, haemostatic or thrombolytic
XX  activity, or anti-inflammatory activity.
XX
XX  Sequence 464 AA:
SQ
XX
XX  Query Match 13.1%; Score 493.5; DB 21; Length 464;
XX  Best Local Similarity 27.6%; Pred. No. 1,le-21;
XX  Matches 142; Conservative 73; Mismatches 194; Indels 106; Gaps 20;
XX
XX  195 ELSSPEYPPYKSLSSCTYSILEBGSYILDVESFDETHPELTCPYDFLKIQTDRRE 254
XX  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX  28 QLTSPGYRPPYKGGSGSDIKAPGFAVRLVF-QDFLE--PSQPCADSVTISFVGS 84
XX  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX  255 HGFPGCK-----TLPHRE--TKSNVTITFTVDESQDHTGKIHVYTAQPCRPMPAP 306
XX  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX  85 PSQFCGQGSPLGRPPGQREYVSGRSRLTLRTTOPSSSENKTAHLH----- 130

```

```

QY  307 PNGHVSPOAKYILKDSFISFCETGVELLQGHLPKLSFTAWCQKDSMDRPMAPACSIYDC 366
DB  131 -----KGFALXYQTVAV-----NSQPISEA----- 151
QY  367 GPPDDLPSGRVEYITGPGVTTTKAVIYQSCETFTTMYKNDCKYVCADGFMTSSKGEKS 426
DB  152 -----SRGSEAIAPAPGDNPAK--VQNHQCEPPYV-QAAMAGALTCAPGTWKDRDCEE 201
QY  427 LPVCEPVCG-----LSARTGGRIYGOKAKRGPDPMPQVLLLGTTAGALLLYDQWVLT 480
DB  202 VLQCMPCVGRPVTPIAQNOQT-----LGSSRAKLGNPMPWAFSTIHGRGGGALLGRWILT 257
QY  481 AAHAYEECK-----HDSALDIRMG-----TLKRLSPHYTQANSEAVFIHEGYTHDA--GF 529
DB  258 AAHTYVPRKDSVSLRKQNSVNFGLHTAIDEMKLGNHP-----HRVYVPRDVRQNSHNF 313
QY  530 DNDIALIKLNKVVINSITPTICLPKKEASTMRDDITAGSGWGLTQRFGLARNLMYVD 589
DB  314 SCDIALLELQHSIPLGPNVLPVCLP--DNETLYRSGLLGVSFGFM-EMGWLTTLELKYSR 370
QY  590 IPVVDHOKCTAYEKPPYRGSVVTANMLCAGLESQKSDCRDGGGALVFLDSETERFV 649
DB  371 LPVAPREACNAWLQKRQPE-VESDMFCVGETQRHSHVCGQSGSVYVWMDNHAHHVA 429
QY  650 GGIVSGSMNCGEAGQGYTRVINYIPWENIIS 684
DB  430 TGIYSWG-IGCGEG--YDFYTKVLSYVDIKGVNM 461
XX
XX  RESULT 16
XX  AAU08684
XX  ID AAU08684 standard; Protein; 487 AA.
XX
XX  AAU08684:
XX
XX  18-DEC-2001 (first entry)
XX
XX  Human FCTR5b polypeptide sequence.
XX
XX  Human; FCTR5b; myelogenous leukaemia; carcinoma; melanoma; glioma;
XX  astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;
XX  neurological disorder; neurodegenerative disorders; nerve trauma;
XX  familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
XX  demyelinating Gardner syndrome; familial myelodysplastic syndrome;
XX  mental health condition; immunological disorder; allergy; infertility;
XX  bronchial asthma; Avelino type eosinophilia; lung disease; deafness;
XX  reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
XX  desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
XX  gastric disorders; pancreatic disease; Schistosoma mansoni infection;
XX  Spino cerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;
XX  Corneal dystrophy-Greeneow type I; Corneal dystrophy-lattice type I;
XX  Reis-Bucklers corneal dystrophy; cystostatic; immunosuppressive;
XX  antidiabetic; antiasthmatic; antileishmanial; antineoplastic;
XX  antidiabetic; antiasthmatic; antileishmanial; antineoplastic;
XX  gynaecological; antifertility; immunostimulant; auditory; haemostatic;
XX  gene therapy; FCTR5b; human complement C1R component precursor.
XX
XX  Homo sapiens.
XX
XX  WO200166747-A2.
XX
XX  13-SEP-2001.
XX
XX  05-MAR-2001; 2001WO-US07160.
XX
XX  03-MAR-2000; 2000US-186592P.
XX  03-MAR-2000; 2000US-186718P.
XX  06-MAR-2000; 2000US-187293P.
XX  06-MAR-2000; 2000US-187294P.
XX  17-MAR-2000; 2000US-190400P.
XX  07-APR-2000; 2000US-196018P.
XX  03-JAN-2001; 2001US-259548P.
XX

```



```

Db 412 ERSQFVVTNSNKTIVRFHSDSYDTGTGLAEYLSDSSDPCP-----GQFTCRTRGC 464
QY 319 ILK-----DSFSICETGYELLQGLPLKSFYAVCQK-----DGSMDRPM 359
Db 465 IRKELRCGDGADCTDHSDELNCSCDAGHQFTCKNKRCKPLFWCDSVNCGSDSDQ 521
QY 360 ACSIV-----DCGP-PDDLPSGRVEYITGGVTTKRAVIQYSC 396
Db 522 GCSCPQTFRCNSNGKCLSKSQCCNGKDDCGSDGDEASCPRVNVVTCRKH-----YRC 574
QY 397 EEFTFMKVN---DGKYVEADGFWTSSKGEKSLPCEPVCGLSARTTGGRIYGGOKAP 453
Db 575 LNCGLSKGNECDGKEDC-SDG-----SDEKD---CD--CGLRSTFRQARVVGTDAD 623
QY 454 GDPFMOVL--LG-GTTAAGALLYDNMVLTAHAAYEOK-----HD---491
Db 624 GEPFMOVSLHALGQGHICGASLSPKMLVSAACHYIDDRGFRYSDPQTQWTFALGLHDQ 663
QY 492 ASALDIRMGTLRKLSPHYTQAMSEAVFIHEGYTHDAGFNDIALIKLNKVVINSNTPI 551
Db 684 RSAPGVQERRLRKRIISH-----PFNFDFDYDIALLELEKPAEYSSWVRPI 730
QY 552 CLPRKAESEFMRTDDIGTASGMLTQGRFL-ARNLMYVDPIYDHQCKTAAYEKPPYRG 610
Db 731 CLP--DASHVFPAGKAIWVTGWGHTGYGTGALLQGEIRVINYOTTENLLPQ-----782
QY 611 SYTNMMLCAGLESGKDSKRGDSGALVFLADSETERMFVGGIYVSGMNGEAGQYGVY 670
Db 783 QITPRMVCVFLSGVDSGCGDSGGLSVYAD-GRIFQAGVSWKD-GCAQNNKRGVYT 840
QY 671 KVINIYPIWE 680
Db 841 RLPLFRDMIK 850

RESULT 19
AAE06930
ID AAE06930 standard; Protein; 855 AA.
XX
AC AAE06930;
XX
DT 16-OCT-2001 (first entry)
DE Human membrane-type serine protease (MTSP) 1.
XX
KM Human: transmembrane serine protease; membrane-type serine protease;
MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
matrilipase.
XX
Homo sapiens.
XX
Key Location/Qualifiers
FT Domain 615..855
FT /label= Protease_domain
XX
PN WO200157194-A2.
XX
PD 09-AUG-2001.
XX
PE 02-FEB-2001; 2001WO-US03471.
XX
PR 03-FEB-2000; 2000US-0179982.
PR 18-FEB-2000; 2000US-0183542.
PR 22-JUN-2000; 2000US-0213124.
PR 26-JUL-2000; 2000US-0220970.
PR 08-SEP-2000; 2000US-0657986.
PR 22-SEP-2000; 2000US-0234840.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Ong EO, Yeh J;
XX

```

```

DR WPI; 2001-488877/53.
DR N-PSDB; AAD13113.
XX
PT Novel single chain polypeptide comprising protease domain of type-II
PT membrane-type serine protease or its catalytically active portion
PT useful for treating and preventing cancer and tumor
XX
XX
Claim 12; Page 195-197; 256pp; English.
XX
CC The invention relates to transmembrane serine proteases and their
CC corresponding nucleotides and the protease domain of a type-II
CC membrane-type serine protease (MTSP). MTSP is useful for identifying
CC compounds that modulate or inhibit its proteolytic activity and for
CC formulating a medicament for treating neoplastic disease. MTSP and
CC its corresponding nucleotides are useful in preventing or treating
CC tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC is useful as a diagnostic marker for tumour development, growth and/or
CC progression and as immunogens to generate antibodies that specifically
CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC gene therapy. The present sequence is human MTSP1 protein (also called
CC matrilipase).
XX
SQ Sequence 855 AA:
XX
Query Match 13.0%; Score 490; DB 22; Length 855;
Best local similarity 25.5%; Pred. No. 3.3e-21;
Matches 186; Conservative 90; Mismatches 270; Indels 184; Gaps 31;
QY 29 RLASGCP--GEVANDERRWTLTAPPGYRLRYFHDELHSLCEYDFVKKSSGAK---84
Db 227 RFTTPGFPDPSYPANARCOMALRGDADSVLSLTFRSFLASCDERGSDDVLYTYNTLSPME 286
QY 85 --VLATLGOESTDTERARAGKDFEYSLGSSLDITFRSDSNKKEPFGFEA-PYAAEDIDE 141
Db 287 PHALVQDC-----TTPSYNTLFHSSQVNLTLITLITERNRP--GEFATFFQLPRMSS 339
QY 142 CQVAPGEAPRTCDHHCHNHLLGFGYCSRAGYVLRNKRKTSALCSGQVFORSGELSPPEY 201
Db 340 C-----GGRLLAKAGTFNSPIY 356
QY 202 PRPYKRLSSCTYTSISLEEGESVILDFVESFDVETH-PEFLCPYDLKIQTDREHGPFCG 260
Db 357 PGHYPPNICTWNIIEVPPNQHVKVSFPHYLLPEVPAGCTCKDYVEINGER-----YCG 411
QY 261 KTLPHLEIKSNVTITFTYDSEGDHTGKIHYS--TAQPCPYPARPNHGVSPYQAY 318
Db 412 ERSQFVVTNSNKTIVRFHSDSYDTGTGLAEYLSDSSDPCP-----GQFTCRTRGC 464
QY 319 ILK-----DSFSICETGYELLQGLPLKSFYAVCQK-----DGSMDRPM 359
Db 465 IRKELRCGDGADCTDHSDELNCSCDAGHQFTCKNKRCKPLFWCDSVNCGSDSDQ 521
QY 360 ACSIV-----DCGP-PDDLPSGRVEYITGGVTTKRAVIQYSC 396
Db 522 GCSCPQTFRCNSNGKCLSKSQCCNGKDDCGSDGDEASCPRVNVVTCRKH-----YRC 574
QY 397 EEFTFMKVN---DGKYVEADGFWTSSKGEKSLPCEPVCGLSARTTGGRIYGGOKAP 453
Db 575 LNCGLSKGNECDGKEDC-SDG-----SDEKD---CD--CGLRSTFRQARVVGTDAD 623
QY 454 GDPFMOVL--LG-GTTAAGALLYDNMVLTAHAAYEOK-----HD---491
Db 624 GEPFMOVSLHALGQGHICGASLSPKMLVSAACHYIDDRGFRYSDPQTQWTFALGLHDQ 683
QY 492 ASALDIRMGTLRKLSPHYTQAMSEAVFIHEGYTHDAGFNDIALIKLNKVVINSNTPI 551
Db 684 RSAPGVQERRLRKRIISH-----PFNFDFDYDIALLELEKPAEYSSWVRPI 730
QY 552 CLPRKAESEFMRTDDIGTASGMLTQGRFL-ARNLMYVDPIYDHQCKTAAYEKPPYRG 610
Db 731 CLP--DASHVFPAGKAIWVTGWGHTGYGTGALLQGEIRVINYOTTENLLPQ-----782

```


QY	611	SVTAMMLAGLSEGGKDCRGDGGALVFLDSETRMFVGGIVSMGSMNCGEAGQGYVT	670
Db	733	QITPRMCCVGLSGGVSDSCQDGGPGLSSVEAD-GRIFQAGVSWGD-GCAQRNKPQYVT	840
QY	671	KVINIPIWIE 680	
Db	841	RLLPFRDMIK 850	
RESULT 20			
ID	AAB98500		
AC	AAB98500 standard; Protein; 855 AA.		
XX			
XX	AAB98500;		
XX			
DT	03-AUG-2001 (first entry)		
XX			
DE	Human TADG-15.		
XX			
XX	Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;		
OS	tumour antigen-derived gene 15; extracellular serine protease.		
XX	Homo sapiens.		
PN	WO200129056-A1.		
XX			
PD	26-APR-2001.		
XX			
PF	20-OCT-2000; 2000WO-US29095.		
XX			
PR	20-OCT-1999; 99US-0421213.		
XX			
PA	(UYAR-) UNIV ARKANSAS.		
PI	O'Brien TJ, Tanimoto H;		
XX			
DR	WPI: 2001-381031/40.		
XX	N-PSDB: AAH23601.		
PT	Novel extracellular serine protease, termed tumor antigen-derived gene		
PT	15 protein overexpressed in carcinomas and DNA encoding it, for		
PT	diagnosis, treatment, prevention of cancer, particularly breast,		
PT	ovarian cancer		
XX			
XX	Claim 11; Fig 2; 130pp; English.		
XX			
XX	The present sequence represents human tumour antigen-derived gene 15		
XX	(TADG-15) protein. TADG-15 is an extracellular serine protease. It was		
XX	found that TADG-15 is over-expressed in ovarian tumours. TADG-15 protein		
XX	or its fragments of 9-20 residues that lack TADG-15 protease activity are		
XX	useful for vaccinating an individual against TADG-15, having, suspected		
XX	of having or at risk of getting cancer. Furthermore, the TADG-15 gene can		
XX	be used as a diagnostic or therapeutic target in cancer.		
XX			
XX	Sequence 855 AA;		
QY	Query Match	13.0%; Score 490; DB 22; Length 855;	
Db	Best Local Similarity	25.5%; Pred. No. 3,3e-21;	
XX	Matches 186; Conservative	90; Mismatches 270; Indels 184; Gaps 31.	
QY	29 RLASGFP-GEYANDQERRMTLTADPGYRLRLYFTHFDELHSLCEYDKLSGAK---	84	
Db	227 RTTGTGFPDSPYPAHARQWALRGDADSVLSLTFPSFLDASCDEROSDLVTVYNTLSPWE	286	
QY	85 --VLTATLGGESTDTERAPKGRDFTYSLGSSLDITFRSDYSNKRKPTGFPA-FYAADIDE	141	
Db	287 PHALVOLCG----TYPSPYNLTFRHSSQVNVLLITLTITNRRHP--GFEATPFOQLPRMS	339	
QY	142 CVAGGEAPTCDDHCHNHLGCFYCSCRAGYVLRHNRKTCALSCGVFQORSGELSSPEY	201	
Db	340 C-----GGRLRKAGGTNSPY	356	
QY	202 PPRPKLSSCTYSILEEGFSVILDFVSEDEVTH-PEITLCYPDFLKIQTDRREHGPFCG	260	

D6		: : : : : : : : : : : : :	357	PGRHPNIDCTWNLIEVPNNQHWKSEKFFYLELLEGVAPAGCPKDYLINSEK-----YCG	411
OY		261	KTLPRRIETKSNNTVTITFEVDDESGDHGHKLHNYS--TAOPCPYPMAPRNGHVSPYOAKY	318	
D6		412	ERSQGVIVTSNSNKITVRHSQSTYTDGFLAEYLSYDSDPCP-----GQFTCRTRGRC	464	
OY		319	ILK-----DSFISFETGYELLQHLPLKSFYAVCQK----DGSWDRPMR	359	
D6		465	IRKELRCGMADCTDHSDELNCSCDAGHQFCCKNFCKPPLFWCDVSNDGDNDEQ---	521	
OY		360	ACSTV-----DCGP-PRDLPSGREYITGGCVTTTKAVIOLYSC	396	
D6		522	GCSCEPAQTFRCSNGSKCLSCSQCCNGKDDCGGSDSEASCPKNVVTCRKHT-----YRC	574	
OY		397	EETFTYMKVN---DGKYUCEADGFMTSSKGEKSLPVCEPYVCSSARTTGGRTYGQKAKP	453	
D6		575	LNGLCLSKGNPECDCKEKC-SDG-----SDKE---CD--CGLSFTQRARVVGCTADE	623	
OY		454	GDFEPWOVLII--LG-GTTAAGALLVDNMVLTAAHAVYEOK-----HD--	491	
D6		624	GEMPOVSLHALGGCHICGASLSIPNMLVSAAHCYIDRGFRYSDPQMTAFGLHDOQ	683	
OY		422	ASALDIRGTATLKRLSPHTQMSSEAVFHEGYTHDAGEFDNDIALIKLNKKVINSNTPT	551	
D6		664	RSAPVDQERRLRKIISH-----PFENDFTFDYDIALLELEKPAEYSSMWPI	730	
OY		552	CLPREKESFMKRTDDIGTASCMGLTORCFL-ARMLVMDVIPYDHOKCTAAVEKPPYRG	610	
D6		731	CLP--DASHVFPACKAIWTVTGHTQYGGTALLLOQGEIRVINOTTGENLLRP-----	782	
OY		611	SVTANMLCAGLESGBKDCSGDSCGALVFLDSETRMRFVGSIGVSMGNCGEAGOVYVT	670	
D6		783	QITPRMCVGLSGVDSCQDSCGSLISSVEAD-CRIFGAGVSVWDG-GCAQRNRKPGVYT	840	
OY		671	KVINIIPWIE 680	.	
D6		841	RLPFRDMWK 850		
RESULT 21					
AA19551					
ID			AA19551 standard; Protein: 683 AA.		
XX					
AC			AA19551;		
XX					
DT			22-JAN-2001 (first entry)		
XX					
DE			Human matrixinase (truncated form).		
XX					
KM			Matriptase: serine protease; human; breast cancer; pre-malignancy;		
KW			actinic keratosis; leukoplakia; Barrett's epithelium;		
KM			columnar metaplasia; ulcerative colitis; Bowenoid papulosis;		
KM			adenomatous colorectal polyp; Ocular erythroplasia;		
KM			vulvar intraepithelial neoplasia; tumour; metastasis; therapy.		
OS			Homo sapiens.		
XX					
FH			Key	Location/Qualifiers	
FT			Peptide	1..31	
FT				/label- signal_peptide	
FT			Protein	32..683	
FT				/label- Mature_protein	
FT			Region	42..155	
FT				/note= "complement subcomponent 1r and 1s (Clr/s) region"	
FT			Region	168..268	
FT				/note= "complement subcomponent 1r and 1s (Clr/s) region"	
FT			Region	280..314	
FT				/note= "LDL-receptor type region"	
FT			Region	315..351	
FT				/note= "LDL-receptor type region"	

FT Region 352..387
 FT /note= "LDL-receptor type region"
 FT Region 394..430
 FT /note= "LDL-receptor type region"
 FT Domain 432..683
 FT /label= Catalytic_domain
 FT Modified-site 130
 FT /note= "N-glycosylated"
 FT Modified-site 313
 FT /note= "N-glycosylated"
 FT Active-site 484
 FT /note= "Forms catalytic triad with Asp-539 and Ser-633"
 FT Active-site 539
 FT /note= "Forms catalytic triad with His-484 and Ser-633"
 FT Active-site 633
 FT /note= "Forms catalytic triad with His-484 and Asp-539"
 FT Peptide 77..79
 FT /note= "RGD motif"
 FT Disulfide-bond 459..485
 FT Disulfide-bond 604..618
 FT Disulfide-bond 629..658
 FT Cleavage-site 442..446
 FT /note= "conserved proteolytic activation site"
 PN WO200053232-A1.
 PD 14-SEP-2000.
 XX 10-MAR-2000; 2000WO-US6111.
 PF 12-MAR-1999; 99US-0124006.
 XX (GEO) UNIV GEORGETOWN.
 PA
 XX
 PI Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;
 DR WPI: 2000-594268/56.
 DR N-PSDB; AAA88492.
 PT Treating malignancies, premalignant and pathologic conditions in a
 XX Subject, comprises administering matrilase modulating agent -
 PS Claim 14; Fig 9; 116pp; English.
 XX
 The present sequence is that of the truncated form of human
 matrilase, a trypsin-like protease, as deduced from cDNA (see
 AAA88492) obtained from human breast cancer cell cDNA by RT-PCR.
 The full-length form (see AAB19552) has an additional 172 N-terminal
 amino acids. Either form can be produced in transformed or
 transfected cells using a claimed method. The zymogen (inactive)
 form of matrilase is a single-chain protein. The active (2-chain
 form strongly interacts with fragments of a Kunitz-type serine
 protease inhibitor (hepatocyte growth factor activator inhibitor,
 HAI-1) to form SDS-stable complexes. In breast cancer cells,
 matrilase is present mainly as the uncomplexed form. Only the
 complexed matrilase is detected in human milk. The invention is
 directed to a method of detecting a malignancy or a pre-malignant
 lesion in breast or other tissue by detecting the presence of
 single- or 2-chain forms of matrilase in the tissue. The
 object is to inhibit tumour onset, tumour growth and metastasis.
 CC Malignancies and pre-malignant conditions characterised by
 CC expression of the zymogen or activated form of matrilase are
 CC treated by administering an inhibitor of matrilase, especially a
 CC Bowman-Birk inhibitor. The pre-malignant condition is atypical
 CC ductal hyperplasia of the breast, actinic keratosis, leukoplakia,
 CC Barrett's epithelium of the esophagus, ulcerative colitis,
 CC adenomatous colorectal polyps, erythroplasia of the Queyret,
 CC Bowen's disease, Bowenoid papulosis, vulvar intraepithelial
 CC neoplasia or dysplastic changes to the cervix. The invention also
 CC provides methods for in vivo or in vitro diagnosis of malignancy

CC or pre-malignant lesion, and methods of identifying matrilase
 CC modulators, including activators and inhibitors.
 XX
 SO Sequence 683 AA:
 Query Match 13.0%; Score 488; DB 21; Length 683;
 Best Local Similarity 25.5%; Pred. No. 3.4e-21;
 Matches 186; Conservative 90; Mismatches 270; Indels 184; Gaps 31;
 QY 29 RLASPGEP-GEYANDGERRWTLAPPGYRLRLYFHFQDLHLSCYDYDKLSSGAK----84
 DB 55 RFTTGPFPSPYPANARCCWALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPME 114
 QY 85 --VLATLCGEESTDTERADGKDTFYSLGSSLDITFSSDSNKEPFTGFEA-FYAAEDIDE 141
 DB 115 PHLVQLCG-----TTPPSNLTFHSSQVLLITLTINERRRP--GFATPFQULPRMS 167
 QY 142 QYVAPGEAPTCDHCHNHLGEGYCSGRAGYVLRNKRKTSALCSGOVFPQSGELSSPEY 201
 DB 168 C-----GRLRKAOGTFNSPY 184
 QY 202 PREPKLSCTYSISLEEGSVILDFVESFDVETH--PETLCYDFLKIQTDRREHPPCG 260
 DB 185 PGHYPRNIDCTWNEVPNNQHVAKRFYLLPEGPVPACTCPDYVEINERK----YCG 239
 QY 261 KTLPHRIETKSNVTITTFVYDESBDHTGKIHYS--TAQPCPYPMAPNGHVSPOQARY 318
 DB 240 ERSGFVYTSNSKTIYRFHSDQSYTDTGFLAEYLSLSDSDPC-----GGFTCTGTGC 292
 QY 319 ILK-----DSFSIFCEYELLOGHLPLKSTFAYCQK-----DGSMDRPM 359
 DB 293 IRRELRCDDGWADCTDHSDELNCSDAGHOFTCKNKFCKPLFWYCDVNDGDNSDQ--- 349
 QY 360 ACSIV-----DCGP--PDDLPSRVEITIGPGVTTKAVIQYSC 396
 DB 350 GCSCPQTERCSNGKCLSSQCCNGKDCGDSDEASCPVNVVYCTKHT-----YRC 402
 QY 397 EEFYFMKVN---DGRVYGEADGFWTSSKGEKSLPYCEPVCISARTTGGRIYGGQAKP 453
 DB 403 LNLCLSKNGPECDKEDC-SDG-----SDEKD---CD--CGLRSTFRQARVYCGTDAE 451
 QY 454 GDFPQVOLI--IG-GTTAAGALLYDMVLTAAHAYEQK-----HD-----491
 DB 452 GEMPQVQSLHALQGHICGASLISPMWLVSAAHCYIDDRGFRYSDEPTQWTAFLGLHDSQ 511
 QY 492 ASALDIRMGTLKRSLPHYQGAMSEAVFIHGGYTHDAGFPNDALILKKNVYINSITPI 551
 DB 512 RSAPGVQERRLKRITSH-----PFPNDTFPDIALLELEKPAEYSSWVRPI 558
 QY 552 CLPRKEAESFMRTDIDGTASGMGLTORGFL-ARNLMYVDIPYDHQCTAAVEKPPYPRG 610
 DB 559 CLP--DASHFPAGKALIVTGMWGHITGYGGALILDKGELRVINQTCENLLPQ-----610
 QY 611 SVTANMLCAGLESGGDCSGDSCGALVPLDSETEWVGVGIVSGSMNGEAGQGYVT 670
 DB 611 QIRPRMKCVGFLSGVDSGDSGGLPSVEAD-GRIFQAGVSMGD--GCAQRNKGYYT 668
 QY 671 KVINIYPIWE 680
 DB 669 RLPLFRDWIK 678
 RESULT 22
 AAY90284
 ID AAY90284 standard; Protein; 762 AA.
 XX
 AC AAY90284;
 XX
 DT 24-OCT-2000 (first entry)
 XX
 DE Human peptidase, HPEP-1 protein sequence.
 XX
 KW Human; peptidase; cell proliferative disorder; arteriosclerosis;

/note="conserved proteolytic activation site"
 FT XX WO200053232-A1.
 PN XX 14-SEP-2000.
 PD XX 10-MAR-2000; 2000MO-US06111.
 PF XX 12-MAR-1999; 99US-0124006.
 PR XX (GEOU) UNIV GEORGETOWN.
 PA XX Dickson RB, Lin C, Johnson M, Wang S, Eneyedy I;
 PI XX WPI; 2000-594268/56.
 DR XX N-PSDB; AAA8493.
 PT Treating malignancies, premalignant and pathologic conditions in a
 PT Subject, comprises administering matrilipase modulating agent in a
 PT Claim 14; Fig 16; 116pp; English.

The present sequence is that of the full-length form of human
 CC matrilipase, a trypsin-like protease. This has an additional 172
 CC amino acids compared with the truncated form of matrilipase given
 CC in AAB19551. Either form can be produced in transformed or
 CC transfected cells using a claimed method. The zymogen (inactive)
 CC form of matrilipase is a single-chain protein. The active 2-chain
 CC form strongly interacts with fragments of a Kunitz-type serine
 CC protease inhibitor (hepatocyte growth factor activator inhibitor,
 CC HAI-1) to form SDS-stable complexes. In breast cancer cells,
 CC matrilipase is present mainly as the uncomplexed form. Only the
 CC complexed matrilipase is detected in human milk. The invention is
 CC directed to a method of detecting a malignancy or a pre-malignant
 CC lesion in breast or other tissue by detecting the presence of
 CC single- or 2-chain forms of matrilipase in the tissue. The
 CC object is to inhibit tumour onset, tumour growth and metastasis.
 CC Malignancies and pre-malignant conditions characterised by
 CC expression of the zymogen or activated form of matrilipase are
 CC treated by administering an inhibitor of matrilipase, especially a
 CC Bowman-Birk inhibitor. The pre-malignant condition is atypical
 CC ductal hyperplasia of the breast, actinic keratosis, leukoplakia,
 CC Barrett's epitheliellum of the esophagus, ulcerative colitis,
 CC adenomatous colorectal polyps, erythroplasia of the Queyrat,
 CC Bowen's disease, Bowenoid papulosis, vulvar intraepithelial
 CC neoplasia or dysplastic changes to the cervix. The invention also
 CC provides methods for in vivo or in vitro diagnosis of malignancy
 CC or pre-malignant lesion, and methods of identifying matrilipase
 CC modulators, including activators and inhibitors.

Sequence 855 AA:

Query Match 13.0%; Score 488; DB 21; Length 855;
 Best Local Similarity 25.5%; Pred. No. 4.3e-21;
 Matches 186; Conservative 90; Mismatches 270; Indels 184; Gaps 31;

QY 29 RLASGPF-GEYANDQERFWTLTAPGYSRLTYFTTHDLESHLCEYDFKLSGAK--- 84
 DB 227 RFTTGGFPDPRYRAHACQALRGDADSVLSLTFRSFDLASCDSRGSDLYVTYVTLSPME 286
 QY 85 --VLAATCGEESTDTERAPGKDFYYSLGSSLDITFRSDYSENEKPFETGFEA-FYAATEDIDE 141
 DB 287 PHALVQCG-----TTPPSYNLTFHSSQNVLLITLITERRHP--GFEATFFQLPRMSS 339
 QY 142 CQVAPGAPFCDHCHNHLGFGYCSCAGYVLRNKRKTCALCSGOVFTQRSGLSPEY 201
 DB 340 C-----GGRLLRAQCGTFNSPY 356
 QY 202 PRPYKLSCSTYSLSIEGFSVILDEVESPDVETH-PELTCPPYFLKIQTDREHGFPCG 260
 DB 357 PGHTPPNIDCTWNEVNNHVKYRFKFFYLLEGVAGCPCPKDYVINGEK-----YCG 411
 QY 261 KTLPHRIETSNVTYITFVDESQDHTGKRIHTYS--TAQPCPYPMAPRNGHVSVPQAKY 318

DB 412 ERSQFVVTSMNSKRTVFRHSDQSYTDTGFLAEYLSYDSSDPCP-----GQFTCRTRGC 464
 QY 319 ILK-----DSISICEFYELLOGHLLDLKSTANVCQK-----DGSMBRPMW 359
 DB 465 IRRELRCGDMACTDHSDELINCSGDAGHOFTCKNFKCKPLFWCVDSVNDGNSDQ--- 521
 QY 360 ACSIV-----DCGP-PDDLPSGVEYITGPGVTTYKAVIQYSC 396
 DB 522 GCSCPAPQTFRCNSGKCKLSKQGCNGKDDCGSDGSDASCPKVVNVYTCMKH-----YRC 574
 QY 397 EETFYTMKVN---DGKYCEADGFWTSSKGEKSLPYCEPYCGLSARTGGRTYCGKAKP 453
 DB 575 LNCGLCKSGNPECDGKEDC-SDG-----SDEKD---CD--CGLRSEFQARVVGTDAD 623
 QY 454 GDFPWOYLI--LG-GTTAAGALLYDMNVLTFAAHAVYEOK-----HD--- 491
 DB 624 GEMWQVSLHALGQGHICGLISPMWLVSAHCYIDDRGFRISDPTQWTAFLGLHDQSO 683
 QY 492 ASALDIRMGLKRLSPHYTQAMSEAVFIHEGYTHDAGFDMDIALIKLNKRVINSNITPT 551
 DB 684 RSAPGVGERLKRILSH-----PFENDFTFDYDIALLELEKPAEYSSWVRPI 730
 QY 552 CLPRKEAESFMRTDIDGTASGGLTORGL-ARNLMYVDIPYDHOKTAAYKPPYPRG 610
 DB 731 CLP--DASHVFPAKAIWYGMGHTQYGGTGAALILQGEIRVINOFTCEMLLPQ----- 782
 QY 611 SYVANNMCAGLESGDSCDGSAGALVFLDSETERFVSGIVSMGNGEAGGYVY 670
 DB 783 QITPRMVCVFLSGVDSCGDSGGLSVLEAD-GRIFQAGVSWBD-GCAQRNKKREYVT 840
 QY 671 KVINYIPWIE 680
 DB 841 RLPLFRMWIK 850

RESULT 24
 AAB35465
 ID AAB35465 standard; Protein; 855 AA.
 AC AAB35465;
 DT 06-JUN-2001 (first entry)
 XX Human membrane-type serine protease MT-SPI.
 DE Human membrane-type serine protease; MT-SPI; cancer.
 KW Human; membrane-type serine protease; MT-SPI; cancer.
 OS Homo sapiens.
 PN WO200123524-A2.
 PD 05-APR-2001.
 PF 02-OCT-2000; 2000MO-US27250.
 PR 30-SEP-1999; 99US-0410362.
 PA (REGC) UNIV CALIFORNIA.
 PI Craik CS, Takeuchi T, Shuman M;
 DR WPI; 2001-245002/25.
 DR N-PSDB; AAF28099.
 PT New nucleic acid encoding a membrane type serine protease, useful for
 PT the diagnosis, prognosis and treatment of cancer, particularly
 PT metastatic cancers -
 PS Claim 7; Fig 1; 102pp; English.
 CC The present invention provides the protein and coding sequences for the
 CC novel human membrane-type serine protease MT-SPI. Increased expression of

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 22:20:01 : Search time 15 Seconds
(without alignments)
922.832 Million cell updates/sec

Title: US-09-874-198-2
Perfect score: 3758
Sequence: 1 MRLTLGLGCGSVATPLGP.....GYTRVINYPIWENTISDF 686

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 75 summaries

Database :
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCT05_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3758	100.0	686	10	US-09-874-198-2
2	3758	100.0	686	10	US-09-874-238-2
3	1531	40.7	679	10	US-09-874-198-6
4	1531	40.7	679	10	US-09-874-238-6
5	1412	37.6	705	9	US-09-808-602-94
6	1406.5	37.4	688	10	US-09-874-198-7
7	1406.5	37.4	688	10	US-09-874-238-7
8	1238	32.9	760	10	US-09-925-301-1024
9	1234	32.8	673	10	US-09-874-198-8
10	1234	32.8	673	10	US-09-874-238-8
11	852.5	22.7	449	10	US-09-925-302-612
12	681.5	18.1	351	10	US-09-925-302-611
13	493.5	13.1	464	9	US-09-808-602-93
14	493.5	13.1	487	9	US-09-808-602-95
15	492	13.1	452	9	US-09-808-602-21
16	492	13.1	487	9	US-09-808-602-19
17	491.5	13.1	487	9	US-09-808-602-17
18	490	13.0	855	9	US-10-099-700A-2
19	469.5	12.5	855	10	US-09-900-751-2

ALIGNMENTS

20	446.5	11.9	269	10	US-09-981-123-2	Sequence 2, Appl1
21	444	11.8	802	9	US-09-978-295A-169	Sequence 169, App
22	444	11.8	802	9	US-09-878-697-169	Sequence 169, App
23	444	11.8	802	9	US-09-878-192A-169	Sequence 169, App
24	444	11.8	802	9	US-09-999-832A-169	Sequence 169, App
25	444	11.8	802	9	US-09-978-189-169	Sequence 169, App
26	444	11.8	802	10	US-09-888-615-113	Sequence 113, App
27	424.5	11.3	970	10	US-09-888-615-101	Sequence 101, App
28	419	11.1	367	9	US-09-977-577-7	Sequence 7, Appl1
29	414.5	11.0	946	10	US-09-285-385C-19	Sequence 19, Appl1
30	410	10.9	415	10	US-09-118-748-2	Sequence 2, Appl1
31	405	10.8	461	10	US-09-884-901-3	Sequence 3, Appl1
32	403	10.7	329	9	US-09-977-577-9	Sequence 9, Appl1
33	402	10.7	347	9	US-09-977-577-3	Sequence 3, Appl1
34	401.5	10.7	720	9	US-09-892-598-231	Sequence 231, App
35	401.5	10.7	720	9	US-09-899-293A-231	Sequence 231, App
36	401.5	10.7	720	9	US-10-063-547-38	Sequence 38, Appl1
37	401.5	10.7	720	9	US-09-989-735-231	Sequence 231, App
38	401.5	10.7	720	9	US-09-990-444-231	Sequence 231, App
39	401.5	10.7	720	9	US-09-989-730-231	Sequence 231, App
40	401.5	10.7	720	9	US-09-890-436-231	Sequence 231, App
41	401.5	10.7	720	9	US-09-991-181-231	Sequence 231, App
42	401.5	10.7	720	9	US-09-993-687-231	Sequence 231, App
43	401.5	10.7	720	9	US-09-989-734-231	Sequence 231, App
44	401.5	10.7	720	9	US-10-004-551-4	Sequence 4, Appl1
45	401.5	10.7	720	9	US-09-997-653-231	Sequence 231, App
46	401.5	10.7	720	9	US-10-174-590-170	Sequence 170, App
47	401.5	10.7	720	9	US-10-176-758-170	Sequence 170, App
48	401.5	10.7	720	9	US-10-063-616-38	Sequence 38, Appl1
49	401.5	10.7	720	9	US-10-175-737-170	Sequence 170, App
50	401.5	10.7	720	10	US-09-989-722-231	Sequence 231, App
51	401.5	10.7	720	10	US-09-989-722-231	Sequence 231, App
52	401.5	10.7	720	10	US-09-989-727-231	Sequence 231, App
53	401.5	10.7	720	10	US-09-989-727-231	Sequence 231, App
54	401.5	10.7	720	10	US-09-989-731-231	Sequence 231, App
55	401.5	10.7	720	10	US-09-989-732-231	Sequence 231, App
56	401.5	10.7	720	10	US-09-991-073-231	Sequence 231, App
57	401.5	10.7	720	10	US-09-990-442-231	Sequence 231, App
58	401.5	10.7	720	10	US-09-991-163-231	Sequence 231, App
59	401.5	10.7	720	10	US-09-993-604-231	Sequence 231, App
60	401.5	10.7	720	10	US-09-990-456-231	Sequence 231, App
61	401.5	10.7	720	10	US-09-989-721-231	Sequence 231, App
62	401.5	10.7	720	12	US-10-006-867-38	Sequence 38, Appl1
63	401.5	10.7	720	12	US-10-052-586-170	Sequence 170, App
64	401	10.7	347	9	US-09-977-577-4	Sequence 4, Appl1
65	401	10.7	1013	10	US-09-942-366-5	Sequence 5, Appl1
66	401	10.7	1013	10	US-09-285-385C-20	Sequence 20, Appl1
67	399.5	10.6	406	9	US-09-977-577-2	Sequence 2, Appl1
68	398.5	10.6	352	9	US-09-808-602-105	Sequence 105, App
69	398.5	10.6	352	10	US-09-888-615-95	Sequence 95, Appl1
70	397	10.6	347	9	US-09-977-577-5	Sequence 5, Appl1
71	396.5	10.6	346	9	US-09-977-577-8	Sequence 8, Appl1
72	396.5	10.6	986	10	US-09-850-048A-4	Sequence 4, Appl1
73	396.5	10.6	1015	10	US-09-285-385C-2	Sequence 2, Appl1
74	395.5	10.5	850	10	US-09-888-615-108	Sequence 108, App
75	395	10.5	1012	10	US-09-285-385C-4	Sequence 4, Appl1

RESULT 1
US-09-874-198-2
; Sequence 2, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jenssensius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874, 198
; CURRENT FILING DATE: 2001-06-04

```

:
:
: PRIOR APPLICATION NUMBER: 09/054,218
: PRIOR FILING DATE: 1998-04-02
: PRIOR APPLICATION NUMBER: 60/042,678
: PRIOR FILING DATE: 1997-04-03
:
: NUMBER OF SEQ ID NOS: 8
:
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO: 2
:
: LENGTH: 686
:
: TYPE: prt
:
: ORGANISM: Homo sapiens
:
: OS-09-874-198-2

```

Query Match	100.0%	Score 3758	DB 10	Length 686
Best Local Similarity	100.0%	Pred. No. 8.2e-276		
Matches 686	0	Mismatches 0	Indels 0	Gaps 0

QY	1	MRLLTLGLLGGASVATPLGPGWMPPEVGRGLASPGPEEVYANDDERMTTLTAPPGYRLRLY	60
Dh	1	MRLLTLGLLGGASVATPLGPGWMPPEVGRGLASPGPEEVYANDDERMTTLTAPPGYRLRLY	60
QY	121	YSNKKPFLGFAFAAADIIDECQVAPBEAPPCDHCHNHLAGCYSCORACGYVLRHKRRC	180
Dh	121	YSNKKPFLGFAFAAADIIDECQVAPBEAPPCDHCHNHLAGCYSCORACGYVLRHKRRC	180
QY	181	SALCSGOVFTQSRGSELSSPEYPRRYPKLSSCTQYSISLSEBFSYILDFVESFDETHPETL	240
Dh	181	SALCSGOVFTQSRGSELSSPEYPRRYPKLSSCTQYSISLSEBFSYILDFVESFDETHPETL	240
QY	241	CPYDFLKIQTDREHGHGPGCKTLPHRIETNSNYLTITFVTDSEGDHTGKIHITSTAQPC	300
Dh	241	CPYDFLKIQTDREHGHGPGCKTLPHRIETNSNYLTITFVTDSEGDHTGKIHITSTAQPC	300
QY	301	PYPAPAPGHHVSPQAKYIILKDSFICFETGYELLQHLPLKSPFAVCKDGSMDRMPA	360
Dh	301	PYPAPAPGHHVSPQAKYIILKDSFICFETGYELLQHLPLKSPFAVCKDGSMDRMPA	360
QY	361	CSIVDCGPPDDLPSGRVEYITGPVTTYKAVIOYSCBEFFYTKMKVNDGKVCCEADGFWTS	420
Dh	361	CSIVDCGPPDDLPSGRVEYITGPVTTYKAVIOYSCBEFFYTKMKVNDGKVCCEADGFWTS	420
QY	421	SKGKSLPVCPCPVCGLSARTTGGITVGGOKAKRQDPFWQVLLIGTTAAQALLYDMWVLT	480
Dh	421	SKGKSLPVCPCPVCGLSARTTGGITVGGOKAKRQDPFWQVLLIGTTAAQALLYDMWVLT	480
QY	541	KVYINSNTITPCLPRKAESFMRTDIDTASGWCLOQRGFLARNLNVIPIDVHOKCTA	600
Dh	541	KVYINSNTITPCLPRKAESFMRTDIDTASGWCLOQRGFLARNLNVIPIDVHOKCTA	600
QY	601	AYEKPPYPRGCVTNNMLCAGLESCKGSCGSDSGALVFLDSETERFVYGGIYSWGSMMC	660
Dh	601	AYEKPPYPRGCVTNNMLCAGLESCKGSCGSDSGALVFLDSETERFVYGGIYSWGSMMC	660
QY	661	GEAGQGYGYTKVINYIPMIENIISDF 686	
Dh	661	GEAGQGYGYTKVINYIPMIENIISDF 686	
RESULT 2			
US-09-874-238-2			
: Sequence 2, Application US/09874238			
: Patent No. US20020082209A1			
: GENERAL INFORMATION:			
: APPLICANT: Jentsenius, Jens Chr.			
: APPLICANT: Thiel, Steffen			
: TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND			

```

: TITLE OF INVENTION:  USES FOR IT
: FILE REFERENCE: 09011-002003
: CURRENT APPLICATION NUMBER: US/09/874,238
: CURRENT FILING DATE: 2001-06-04
: PRIOR APPLICATION NUMBER: 09/0054,218
: PRIOR FILING DATE: 1998-04-02
: PRIOR APPLICATION NUMBER: 60/042,678
: PRIOR FILING DATE: 1997-04-03
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 686
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-874-238-2

```

Query Match	100.0%	Score 3758	DB 10	Length 686
Best Local Similarity	100.0%	Pred. No. 8	2e-276	
Matches 686	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	MRITLILGLCGSVATPLGPRWPERPVGRLASPEEPGEVANDORRMTITAPPEYRLRLY	60
Db	1	MRLLTLILGLCGSVATPLGPRWPERPVGRLASPEEPGEVANDORRMTITAPPEYRLRLY	60
QY	61	FTFHDELESHLCEYDFVKLSGKAVLTLTGQESTEDTERAKGKTFYSLSSLDITFRSD	120
Db	61	FTFHDELESHLCEYDFVKLSGKAVLTLTGQESTEDTERAKGKTFYSLSSLDITFRSD	120
QY	121	YSNEKPFIFGFAFAAEDIDBCQVAPGEAPICDHHCNNHLDGFCSCSRAGYVILHRNRTC	180
Db	121	YSNEKPFIFGFAFAAEDIDBCQVAPGEAPICDHHCNNHLDGFCSCSRAGYVILHRNRTC	180
QY	181	SALSCGOYETRQSELSPEYPRYPRKLSCTSYISLEEGSVTLIDVESPDVETHEPTL	240
Db	181	SALSCGOYETRQSELSPEYPRYPRKLSCTSYISLEEGSVTLIDVESPDVETHEPTL	240
QY	241	CPYDFLKIQTDREBHGFCGKTLPHRIETKSNVTITFVYDESGDHGWKIHNTSTAQC	300
Db	241	CPYDFLKIQTDREBHGFCGKTLPHRIETKSNVTITFVYDESGDHGWKIHNTSTAQC	300
QY	301	PYPAAPRPGHVSPIQAKYILKDSFISCEYGYELLQGHLPKSTFAYCOKDGSMDRMPA	360
Db	301	PYPAAPRPGHVSPIQAKYILKDSFISCEYGYELLQGHLPKSTFAYCOKDGSMDRMPA	360
QY	361	CSIYDCGPRDDLPSGRVEYITGPEVTTYKAVIOYSCETFTYKVNNGKYVCADGFWTS	420
Db	361	CSIYDCGPRDDLPSGRVEYITGPEVTTYKAVIOYSCETFTYKVNNGKYVCADGFWTS	420
QY	421	SKGSESLPVCBPVCGLSARTTGGRIYGGOKAKPDPFMQVLLIGGTTAAGALLVDNNVLT	480
Db	421	SKGSESLPVCBPVCGLSARTTGGRIYGGOKAKPDPFMQVLLIGGTTAAGALLVDNNVLT	480
QY	481	AAHAHVYEOKHDAKSLDIRMGTLKRLSPHYTQAMSEAVFIHEGYTHDXGFPNDIALIKLNN	540
Db	481	AAHAHVYEOKHDAKSLDIRMGTLKRLSPHYTQAMSEAVFIHEGYTHDXGFPNDIALIKLNN	540
QY	541	KVVINSNTTPLCEPRKEAESFMRTDDIGTASGMGLTORGFLARLMLVVDPIYDHOKCTA	600
Db	541	KVVINSNTTPLCEPRKEAESFMRTDDIGTASGMGLTORGFLARLMLVVDPIYDHOKCTA	600
QY	601	AYEKPPYPGVSVTANMLCAGLESGBKDCNRDSSGALVFLDSETERMFVGIVSGMSMC	660
Db	601	AYEKPPYPGVSVTANMLCAGLESGBKDCNRDSSGALVFLDSETERMFVGIVSGMSMC	660
QY	661	GEAGQYGYTKVNTYITWIENTIISDF 686	
Db	661	GEAGQYGYTKVNTYITWIENTIISDF 686	

Db 651 -DCGKDRYGVYXYIHNKDMIQRV 674

RESULT 5
US-09-808-602-94
Sequence 94, Application US/09808602
Patent No. US20020155115A1

GENERAL INFORMATION:
APPLICANT: Vernet, Corine A
APPLICANT: Fernandes, Elma
APPLICANT: Shimkets, Richard A
APPLICANT: Heriman, John L
APPLICANT: Majumder, Kumud
APPLICANT: Mishra, Vishnu
APPLICANT: Mezes, Peter S
APPLICANT: MacDougall, John
TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 15966-697 CIP
CURRENT APPLICATION NUMBER: US/09/808,602
PRIOR APPLICATION NUMBER: 09/800,198
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: 60/186,596
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 94
LENGTH: 705
TYPE: PRT
ORGANISM: Homo sapiens
US-09-808-602-94

Query Match 37.6%; Score 1412; DB 9; Length 705;
Best Local Similarity 41.7%; Pred. No. 8.2e-99;
Matches 301; Conservative 108; Mismatches 248; Indels 64; Gaps 22;

1 MRLITLL-GLLC---GSVATPLGPKWPEPVFGRIASPGFGEYANDQERRMTLAPPGY 55
1 MMLLYLVLPALFCRAGSGI--PI----PQKLFGEVTSPLPKPRYNNFETTVITVYTG 54
56 RLRLYTHFDLSESHLCEYFVKLSGAKVATLCGOESTDTERAPKDTFYSIGSSLDI 115
55 RVKLVFOFLEPSESGCYFVYVKISADKSLGRFCGLGSPPLGNPKKFFMSQGNMML 114
116 TFRSDYENKPR-----FTGEAFYAAEDIDEC--QVAPGE--APTCDHCHNHLGFEYC 165
115 TFRHDFSENEGTIMFKGFLAYIAVADDEBCASRSKSGEEDPOPCOHLCNHTVGGIFC 174
166 SCRAGYVLHNRKRTCSALCSGOVFTORSGLSPPEYRPYKLSCTYSISLEGFSVIL 225
175 SCRGYELDHRHSCAQECSELEYTEASGYISLEYRSPYPPDLRCYSTRVERGGLTLHL 234
226 DPEVSFVENHPELICYDYLKIQTDREHGPFCGKTLPHRIETKSNVTYITVYDES 285
235 KLEPEFDIDHQVHCYDLOIYANGNKNGEFGKORPPDLTSSNAVDLFFTTDES 294
286 HTGKIHYSYTAOCPCRYPMAPRN-GHVSPOAKYILKDSIFCEYELLGHLPLKSF 344
295 SRGKMLYTTETIICPKPRLDETTIIONLOPOQFDFYIATCKOQYOLEGNOVLHSF 354
345 TAVCOXKMDRPMACISVDCGPPDLPSGRVEYITPGVYTYKAVIYSCETFTYTMK 404
355 TAVCOQDGTNHRAMPKIKDCGPRRLPBGDFRYTTTGMVNTYKAIQYCHHPYKMQ 414
405 V-----NDGKYVCEADGW-TSSKGEKSLPCEPVCG--LSARTTGGRIYGGQAKPG 454
415 TRAGSRESEGGVYTCYTAOGIMKNEQKEG--IPRLCPVCGKRPVNEVEORRILIGOKAKMG 473
455 DPEMOYVILGGTTAAGLLLDNNVLTAAHAYEDKHDA---SALDIMG-----TLKRLS 506
474 NPEMOYFTNIGHRGGALLDGRWILTAAHTLYPREHEAQSNAISLAVLGHNTVEELMKLG 533

QY 507 PHYTOAMSEAVFIHEGYTHDA--GFDNDIALIKLNKVINISITPICLPKKEAESFMRT 564
Db 534 NHPILR-----RVSVHPDYRODESYNFEEDIALLENSVTLGRLPLPICLP--DNDFEYDL 587
QY 565 DDIGTASGWLTOGRGLIARLMTVDIPIYDHQKCTAYEKPPYPRGS-----VTANMLCA 619
Db 588 GLMGVYSGFCVMEPEK-IAHLRFVRLPVANPOAC-----ENMLRGKRNMDYFSQNMFC 640
QY 620 GLSEGGKDCRGSGGALVLDSETERFWFGIVSGMSNMCGAAGQGYVTKYINTIPMI 679
Db 641 GHPSLKODACQSDSGGVFAVRDNTDRWATGIVSMG-IGCSNG--YGFYTKVLTVDWI 697
QY 680 E 680
Db 698 K 698

RESULT 6
US-09-874-198-7
Sequence 7, Application US/09874198
Patent No. US20020082208A1
GENERAL INFORMATION:
APPLICANT: Jensenius, Jens Chr.
APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
TITLE OF INVENTION: USES FOR IT
FILE REFERENCE: 09011-002002
CURRENT APPLICATION NUMBER: US/09/874,198
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 688
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-198-7

Query Match 37.4%; Score 1406.5; DB 10; Length 688;
Best Local Similarity 41.9%; Pred. No. 2.1e-98;
Matches 291; Conservative 106; Mismatches 244; Indels 53; Gaps 18;

23 PEVVFGLASPGGPEYANDQERRMTLTAPRGYRLLYTHFDLSESHLCEYDFVLTSSG 82
5 PQKLFGEVTSPLPKPRYNNFETTVITVYTGVRKLVFOQFDEPSEGCYDYVKISAD 64
83 AKVATLCGOESTDTERAPKDTFYSIGSSLDITFRSDYSENEKPR-----FTGEAFYAAE 137
65 KSLGRFCGLGSPPLGNPKKFFMSQGNMMLTFTHDFSENEGTIMFKGFLAYIAV 124
QY 138 DIDEK--QVAPGE--APTCDHCHNHLGFEYCSCRAGYVLHNRKRTCSALCSGOVFTOR 192
Db 125 DDEBCASRSKSGEEDPOPCOHLCNHTVGGIFCSCRPYELDHRHSCAQECSELEYTEA 184
QY 193 SGELSPPEYRPYKLSCTYSISLEGFSVILDFVESPDVENHPELICYDYLKIQTD 252
Db 185 SGYISLEYRSPYPPDLRCYSTRVERGGLTLHLKFLPEPDIHQVHCYDLOIYANG 244
QY 253 EEHGPRCGKTLPHRIETKSNVTYITVYDESQDHTGMKIHYSYTAOCPCRYPMAPRN-GH 311
Db 245 KNIGFCGKRRPPDLTSSNAVDLFFTTDESQDSRGKMLRYTTETIICPKPRLDEFTI 304
QY 312 SPVQAKYILKDSIFCEYELLGHLPLKSFYAVCOXKMDRPMACISVDCGPPDL 371
Db 305 ONLOPOQFDFYIATCKOQYOLEGNOVLHSFTAVCOQDGTNHRAMPKIKIKDCGPRN 364
QY 372 LPSGRVEYITPGVYTYKAVIYSCETFTYTMK-----NDGKYVCEADGW-TSSKG 423
Db 365 LPNGDFRYTTTGMVNTYKAIQYCHHPYKMQTRAGSRESEGGVYTCYTAOGIMKNEQKG 424

[illegible]

```

RESULT 7
US-09-874-238-7
; Sequence 7, Application US/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-007003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-874-238-7

```

Query Match	37.4%	Score 1406.5	DB 10	Length 688
t Local Similarity	41.9%	Pred. No. 2.1e-98		
ches 291	Conservative 106	Mismatches 244	Indels 53	Gaps 18

Db	5	PQKLFGEVTSPLFKPKPRPNNETTVTVTPRGYKVLVFOQFDEBEGCFYDIYKISAD	64
QY	83	AKVLATLQGEISTDTERAPGKDFYSLGSSLDITFRSDYSNEK- - -FTGEAFYAAE	137
Db	65	KKSJGRCFGQGLSGPLGPNPGKKEFMSOGNMILLTFHTDFNSENGTIMPGFELAYYQAV	124
QY	138	DIDEC- -OVARGE- -APTDHCHNHLGFGFCGCRAGYLVLRHKRKISALCSQVFTOR	122
Db	125	DLDGCASRKSSEEDPOPQOCHLCHNVGSGFCSCRGYELQEDHRHSQACESSSETLEA	184
QY	193	SGELSPDEYPRPYKLSCTYSISLEEGFYILDFVEFVEVETHPFLCPYDFLKIOTDR	252
Db	185	SGYISLSLEYRSRYPDLRCNYSIFVEGGLTLHLKFLERFDIDHQVCHPCYDLOLIYANG	244
QY	253	EEHGFECGKTLPHRIETKSNVTITFVYDSEGHTGAKIHYSTAORCPYPMARP- GHV	311
QY	312	SPVOAKTILKSEIFCEYETVELLQGLHLKPSFAVQKQSGMSDRMPACSIYDCGPPD	371
Db	305	QNDLPQOFQRFYATCKQGLLEGGQVLHSAFVAVCDDGTWIRAPRCKIKDCCGPRN	364

Qy	372	LPGRVETLIPRGVTTKAVLOYSCETFTPMK- - - - -	NOCKYUCEDADGN- - - - -	ISSKG	423
Db	365	LPNDQFRTTMTGVTYKARLOYUCHERYUKMOGRAGSRESEOGVYUCTOOGIWKNEOK			424
Qy	424	EXSLPVCERVGC- - -LSARATTCGRUYGOKKAPDGFPMOVLILGTTAAGLLALVDNMV/LTA			481
Db	425	EK- - -IPRCLPVCCKVNPVEOKRILIGOKKAMGNFPMOVFTNINHGKGGALLDGRW/LTA			483
Qy	482	AHAAVEQKHDA- - -SALDIRMG- - - - -	TLKLSPHYUQAMSEAVFIHEGTYHDA- - -GFDN		531
Db	484	AHLYLPKHEQNASASLDVFLGHTNVEELMKLGNHPR- - - - -	KXSNPRDYRQDESYNFECS		539
Qy	532	DIALTKLNKKVYINSNTPICLPKKEABESEMRITDICTASGWLTOGRGLARNLMTYDIP			591
Db	540	DIALTELENSYTLGPNLPICLP- - -DNDTFYDLGLMGVYSOFCEWMEK- - -IAHDLRFVRLP			596
Qy	592	IVDHOQKCAALAEKRPYRGS- - - - -	VJANMICALLESBGKDSGDSGSGALVFLDSTER		646
Db	597	VANPDAC- - - - -	BNWLKGNRMDFVSONNEFCAGHPILKQDACGDSGGSVAFVADRPDTR		650
Qy	647	WFGVGIWMSGMNCGEAGOVVUTTKVINYIPIWIE			680
Db	651	WVATGCIWSMG- -IGCSRG- - -YGFYTKVLYNYDMIK			681

```

RESULT 8
US-09-925-301-1024
Sequence 1024, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1024
LENGTH: 760
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (330)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1024

```

[illegible]

Db 304 GNCDSLIVFVAGDRQFGPGCHGFPGLNLETCKSNALDILFOTDLTGOKKWLRYHGD 363
Qy 298 QPCGYPAAPRGVHSPVOAKYILKDSFICETGYELLQGLPLKSTPAVQKXGMDRP 357
Db 364 MPCG-KEDTPNSVWEPAKAKVFRDVQVITCLDGFVEVGVATSFYSTCQSNKMSNS 422
Qy 358 MPACSIYDCGPPDDLPSGAVREYITGPGVTYKAVIOYSCSEFFTYMK-VNDGKYVCADG 416
Db 423 KLCQAPVDCGIPESIEENGKVE---DPESTLFGSVIRTCCEPPYYMMNGGGEVHCANG 479
Qy 417 FWSSKKEKSLPCEPVCGLSAR--TTGGRIYGGQAKKPGDFPMOVLIGSTTAAGALLY 474
Db 480 SWNEVGLPELIPKCVPGVGPVREPFEKORIGGSDADIKFPMQV-FFDNPMAGALIN 538
Qy 475 DNWVLTAAHAYEOKH-----DASALDIRMGTJLRLSPHTQAMSEAVFIHEGY----- 523
Db 539 EYWLTAHAYEGNRKEPTMYVGTSTVOTSRILAKSKMLTP-----EHVFTHPMKLLLEV 591
Qy 524 -THDAGFNDIALIKLNKVINNSNITPICLPKREASFMRDIDGTASGWLTFORGFELA 582
Db 592 PEGRTNFDNDIALVRLKDPVKMGPTVPICLPQTSSDYNLMDGDLGLISGKRTKRDRA 651
Qy 583 RNLMYVPIPYDHQKC-TAAVEKPPYRGS--VTANMLCAGLESGKDCRDSGALVF 639
Db 652 VRLKAARLPVAPLRCKKEVKEKPTADAEAVFTPNMIGAGGEK-GMDSCKGDSGAFAY 710
Qy 640 LD-SETERMFVGIVSGMNCGEAGOGYVTKYINYPWI 679
Db 711 QDPMDKTKFYAAGLVSWGP-----OCGYTLGTRVKNYVDWI 747

RESULT 9
US-09-874-198-8
; Sequence 8, Application us/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; ID NO 8
; LENGTH: 673
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-874-198-8

Query Match 32.8%; Score 1234; DB 10; Length 673;
Best Local Similarity 38.9%; Pred. No. 2,1e-85;
Matches 264; Conservative 101; Mismatches 272; Indels 42; Gaps 18;
Qy 24 EP-VFGLASPGFPEGEVANDQERRWTLTAPGYRLRYFTFHDLIELSHLCEYFVKLSG 82
Db 1 EPTMYGELISPNYFOAYPSEVEKSMDEIEVEGYIHLYFTHLDELSENCAVDSVQIISG 60
Qy 83 AKVLATLCGQESDTEAPKDKTFYSLGSSLDITFRSDYNEKPFGEAFYAEDIDEC 142
Db 61 DTEBGRICGGRSSNNPSPVLEEFOVPYKLYIFKSDFSNEERFTGFAAYVATDINEC 120
Qy 143 QVABGEAPTDHCHNHILGFGYSCRAGYVLRNKRRTCSALSGOVTOSGELSPREY 202
Db 121 TDVY-DVP-CSHCNNMTIGGTFCCPPEYFLHDMDKMGVNCSDVTALIGELASPNYP 178
Qy 203 RPYVPLSSCTYSISLEGGFVILDF-VESFDVETHPTLCPYDELKIQTDREHGPECGK 261

Db 179 KPYPENSRCVEQJRLKEGFQVAVVTLRREDFDVEADAGNCLDSLVEVAGDRQFGPYCGH 238
Qy 262 TLPH--RIEFTKSNVTYTTPTVDESQDHTGKIHNTSTAQCPRPMAPPNHVSPOAKYI 319
Db 239 GFGPPLIETKSNALDILFOTDLTGOKKWLRYHGDMPGP-KEDTPNSVWEPAKAKYV 297
Qy 320 LKDFSIFCEGYELLQGLPLKSFPAVCOKDSMDRPMACSIYDCGPPDDLPSGAVREY 379
Db 298 FRDVQVITCLDGFVEVGVATSFYSTCQSNKMSNKLKCPQVDCGIPESIEENGKVE- 356
Qy 380 ITGPGVTYKAVIOYSCSEFFTYMK-VNDGKYVCADGFMFTSSKGEKSLPVEPVCGLSA 438
Db 357 --DPESTLFGSVIRTCCEPPYYMMNGGGEVHCANGSWNEVGLPELIPKCVPGCVPR 414
Qy 439 R--TTGGRIYGGQAKKPGDFPMOVLIGSTTAAGALLYDNWVLTAAHAYEOKH----- 490
Db 415 EPEERKORILIGGSDADIKFPMQV-FFDNPMAGALINWVLTAAHAYGNEPTMYVG 473
Qy 491 DASALDIRMGTJLRLSPHTQAMSEAVFIHEGY-----THDAGFNDIALIKLNKVI 544
Db 474 STVQTSRLAKSKMLTP-----EHVFTHPMKLLVEPEGRTNFDNDIALVRLKDPVKM 526
Qy 545 NSNITPICLPKREASFMRDIDGTASGWLTFORGLARNLMYVPIPYDHQKC-TAAVE 603
Db 527 GPVSPICLPQTSSDYNLMDGDLGLISGKRTKRDRAVRLKAARLPVAPLRCKKEVKE 586
Qy 604 KPPYRGS--VTANMLCAGLESGKDCRDSGALVF-LD-SETERMFVGIVSGMNC 660
Db 587 KPTADAEAVFTPNMIGAGGEK-GMDSCKGDSGAFAYQDPMDKTKFYAAGLVSWGP--- 642
Qy 661 GEAGOGYVTKYINYPWI 679
Db 643 -QCGTYGLTRVKNYVDWI 660

RESULT 10
US-09-874-238-8
; Sequence 8, Application us/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 673
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-874-238-8

Query Match 32.8%; Score 1234; DB 10; Length 673;
Best Local Similarity 38.9%; Pred. No. 2,1e-85;
Matches 264; Conservative 101; Mismatches 272; Indels 42; Gaps 18;
Qy 24 EP-VFGLASPGFPEGEVANDQERRWTLTAPGYRLRYFTFHDLIELSHLCEYFVKLSG 82
Db 1 EPTMYGELISPNYFOAYPSEVEKSMDEIEVEGYIHLYFTHLDELSENCAVDSVQIISG 60
Qy 83 AKVLATLCGQESDTEAPKDKTFYSLGSSLDITFRSDYNEKPFGEAFYAEDIDEC 142
Db 61 DTEBGRICGGRSSNNPSPVLEEFOVPYKLYIFKSDFSNEERFTGFAAYVATDINEC 120
Qy 143 QVABGEAPTDHCHNHILGFGYSCRAGYVLRNKRRTCSALSGOVTOSGELSPREY 202

Db 121 TDVF-DVF-CSHFCNNFTGGTFCSCPPREYFLHDDKNCVCNCGDVTALIGETASPNVP 178
Qy 203 RPYKLSSCTYSISLEGGFVILDF-VESFVETHPETLCPDFLKIOTDREEHGFCGK 261
Db 179 KPYENSCEQOINLEKGFQVVTILRRDPVLEADSNCLDSLVPFAGDRQCEPRYCGH 238
Qy 262 TLPH-RLETKSNVTYITFTVDESODHTGKIHVTAOPCPYPMPAPNGHVSPOAKYI 319
Db 229 GFGPLNETIKSNMADLITFOTDLTGOKKMKLRYHGDPMPCP-KEDTNSWNEPAAKAYV 297
Qy 320 LKDSSTFCFCEYELLOGLHLPLKSTFANVCQKDSMDRPMRACSTYDCGPPDDPSGRREY 379
Db 298 FRDVVQITCLDGFEEVEGRVATSEYTSYCSNGKWSKLCQPYDCGIPESIEGKVE- 356
Qy 380 ITGCVTFYKAVIOYSCSEETFEYTMK-VNDGKYVCADGFMTSSKGEKSLVPCPEYCGLSA 438
Db 357 --DDESTLFGSVIRTCGEPRYYMENGSGGEYHCAGNSWVNEVGLPBLPKVCYPCGVP 414
Qy 439 R--TTGRTIYGGQAKPGDFPMQVLLIGGTTAAGALLYDNMVLTAHAHAYEOKH----- 490
Db 415 EPFEKQRIIGSGDADIKNFPQV-FPDNPMAGALINEYWLTAHAYVEGNREPTMYG 473
Qy 491 DASALDIMGTLKRLSPHYTOAMSFAVFIHGGY-----TIDAGDNDIALIKLNKKYI 544
Db 474 STSVQTSRLAKSKMLTP-----EHVFIHPGWKLEVEPGRTNENDIALVRLKDPM 526
Qy 545 NSNITPICLPKREASFMRDIDIGTASGMLTORGFRLANLMYVDIPYDHOKE-TAAYE 603
Db 527 GPTVSPICLPCTSSDYNLMODDLGLISMGRTKREKDRVRLKAALPLPAFLKCKEYAVE 586
Qy 604 KPPYPRGS--VTANMLCAGLESCKGKDSGSGALVFLD-SETERWFGVIGSVNGSMNC 660
Db 587 KPTADAEAYVFTPMNICAGGEK-GMDSCKGDGSAFAVQDPNDKTFYAAAGLVSGP--- 642
Qy 661 GEAGGYCTYTVINVIPIWI 679
Db 643 -QCGTYGLYTRVKNYVDWI 660

RESULT 11
US-09-925-302-612
; Sequence 612, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 612
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (284)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-612

Query Match 22.7%; Score 852.5; DB 10; Length 449;
Best Local Similarity 41.2%; Pred. No. 7,4e-57;
Matches 186; Conservative 63; Mismatches 159; Indels 43; Gaps 15;
Qy 256 GPGCGTLPHRIETKSNVTYITFTVDESODHTGKIHVTAOPCPYPMPAPNGHVS 314

Db 9 GEFCKQRXPPDLDTSSNAVDLFFTEDSGDSRGWKLRYTTETLIKPOPKTIDEFTIIONL 68
Qy 315 QAKYLNKSFSTFCEYELLOGLHLPLKSTFANVCQKDSMDRPMRACSTYDCGPPDDLP 374
Db 69 QPOYFRODYFLATYATCKOGYQLIEGNOVLHSFTAVACODDSTHWRAMPKCIKCGOPRLPN 128
Qy 375 GRVEYITPGVYTYKAVIOYSCSEETFEYTMKV-----NDGKYVCADGFMTSSKGEKS 426
Db 129 GDFRTYTMGVNTYKARLOYICHERPYKMQTRAGSRESEGYTCTAGIINKNEKGEK- 187
Qy 427 LPVCEPYCG--LSARTTGRTIYGGQAKPGDFPMQVLLIGGTTAAGALLYDNMVLTAHA 484
Db 188 IPRCLPVCGRPVNPEVQORILIGOKAKMGNFPMQVFTNHCORGALLGDWILTAHT 247
Qy 485 VYEOKHDA--SALDIRG-----TLKRLSPHYTOAMSFAVFIHGGYTHDA--GDDNDIA 534
Db 248 LYPKEHEAQSNAISLDVFLGHTNVEELMKLGNHPIRRXS-----VHPDYQDSESYNEGDIA 303
Qy 535 LIKLNKVVINSNITPICLPKREASFMRDIDIGTASGMLTORGFRLANLMYVDIPYD 594
Db 304 LLELENSVTILGPNLLPILP--DNDFTYDLGLMGVSGFGWMEK-IAHDLRFVRLPVAN 360
Qy 595 HQKTAAYEKPPYPRGS-----VTANMLCAGLESCKGKDSGSGALVFLDSETERWFGV 649
Db 361 PQAC-----ENMLRGKRNMDVFSQNMFCAGHPSLKQDACQDSGSAFAVRDPNTDRMVA 414
Qy 650 GGIYVSGMNCGEAGQYGVYTKRVINVIPIWI 680
Db 415 TGIYSWG-IGCSRG--YGFYTKVLMYVDMIK 442

RESULT 12
US-09-925-302-611
; Sequence 611, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (335)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-611

Query Match 18.1%; Score 681.5; DB 10; Length 351;
Best Local Similarity 43.0%; Pred. No. 4.3e-44;
Matches 138; Conservative 50; Mismatches 112; Indels 21; Gaps 7;
Qy 1 KMLTLTL--GLLC--GSAVAPLGGPKWDEPVFGRLASGFPGEYANDQERWTLTAPPGY 55
Db 2 MWLTLVLPALFCRAGGSI--PI-----POKLEGEVTSPLFPKPYNNFETTYIVTPGY 55
Qy 56 RLRLFTFTFDELSHUCEDYKSSGAKVLAATLGGQSTDTERRAPGADTFPISSGSSIDI 115
Db 56 RVLKVFQDFLEPSGCEGYDYVYKISADKSLGRCFGOLGSLGNPGRKKEPMSQGNKML 115

Qy	116	TFRSDYSNMRP-----PTGFAFAFPAADDEDIC--QVAPGE----APFDHCHHNLGSGFYC	165
		::: :	
Db	116	TFHDEHSNEEGTLMFXYKGFATAYQAADVDEBCASRSKSGEEDPOPOCHLCHNVGVGYFC	175
		::: :	
Qy	166	SCRAGYVYLHRNRKRTCSALSGOVFTORSGETLSPEYRPYPKLSSCTSYISLEEGFSVIL	225
		::: :	
Db	176	SCRQYELQIEDERHSCQAFCSSELTETSGYISLSEIYRSTPPDRCNYSIRVERGCLTLHL	235
		::: :	
Qy	226	DFVESFDEYTHPEFLCPYDFLKIOTDREEHQPFQCGKTLPHRIETKSNVTWITFEVDESGD	285
		::: :	
Db	236	KFLPEFDIDHDHQVHCPCYDQLOIYVANKNIGEFCCGRPPDLTSSNAVDLLFTFTBESGD	295
		::: :	
Qy	286	HTGKKIHHTSTAOCPPIPMAR	306
		::: :	
Db	296	SRGKKRLRYTETEXHQPAPADP	316
		::: :	

RESULT 13
US-09-808-602-95
Sequence 95, Application US/09808602
Patent No. US20020155115A1

```

APPLICANT: Vernet, Corine A
APPLICANT: Fernandes, Elma
APPLICANT: Shinkets, Richard A
APPLICANT: Herrman, John L
APPLICANT: Majumder, Kumud
APPLICANT: Mishra, Vishnu
APPLICANT: Mezes, Peter S
APPLICANT: MacDougall, John
TITLE OF INVENTION: NO. US20020155115A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 159665-697 CIP
CURRENT APPLICATION NUMBER: 09/808,602
CURRENT FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 09/800,198
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: 60/186,596
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 95
LENGTH: 464
TYPE: PRT
ORGANISM: Homo sapiens
US-09-808-602-95

```

Query Match	13.1%;	Score 493.5;	DB 9;	Length 464;
Best Local Similarly	27.6%;	Pred. No. 9.4e-30;		
Matches 142; Conservative	73;	Mismatches 194;	Indels 106;	Gaps 20;

Dd	195	ELSSBEPYRPYKLSCTSYSLSEGEVLIIDEFESDVEHNEPTLPPYFLKJOTDREE	254
	28	QUTSPGEYEPYKGGQESSTIDIKAEQFAVRVLVF-QDEPDLE--PQDCAGDSVLTISFYGSD	84
Qy	255	HGRPCGK-----ILPHRIE--TFSNVTYITFEVTDSCGDHGTGKINHSTRAQCPPIYMAP	306
Dd	85	PSQCGQGGQSGPLGRPPQOREFVSSGSRSLRTLFRQPSSEKNTAHL-----	130
Qy	307	PNGHVSIPVQAKYIILKDSFSEFCETGEYELLOGHLEILKSFSTAWCKDQSDMPACSTYDC	366
Dd	131	-----KGLLALYQTVA-----NYSQPIISA-----	151
Qy	367	GRPDDELPSGRKEVYITTGCVTTTAAVIOYSCSEETFTYAKVNDGKYVCEADGEFTWSSKGEKS	426
Dd	152	-----SRGSEALINAGDNPAAK--YONHCOEPRY--DAAAAGALTCATPGTKWDRODGE	201
Qy	427	LPVCEPYCG-----LSARTTGRIRYGGOKAKRPDPYMOVYLIGCTAAGALILYDNMVL	480
Dd	202	VLCGMPCVGRVPTPIAQNQTT-----LSSRAKLGNFPFQWAFSTISHGGGAGALLDGRNITL	257
Qy	481	AAAHAVYEDK-----HDSALDIRMG-----TLKRLSPHYTQAMSEAVNTIEGCTYHDA--GF	529
Dd	258	AAAHVYPRDVSILKRNOSVAVNLFHTAIDELKLTGKNRPV-----HRVVVHPDYRONESHNF	313

[illegible]

RESULT 14
US-09-808-602-93
; Sequence 93, Application US/09808602
; Patent No. US2002015115A1
; Journal: TUNCDST001

```

? APPLICANT: Fernel, Corine A
? APPLICANT: Fernandes, Elma
? APPLICANT: Shimkets, Richard A
? APPLICANT: Herrman, John L
? APPLICANT: Majumder, Kumud
? APPLICANT: Mishra, Vishnu
? APPLICANT: Mezes, Peter S
? APPLICANT: MacDougall, John
? TITLE OF INVENTION: NO. US20020155115a1 Proteins and Nuclec Acids Encoding Same
? FILE REFERENCE: 15966-697 CIP
? CURRENT APPLICATION NUMBER: US/09/808,602
? CURRENT FILING DATE: 2001-03-14
? PRIOR APPLICATION NUMBER: 09/800,198
? PRIOR FILING DATE: 2001-03-05
? PRIOR APPLICATION NUMBER: 60/186,596
? PRIOR FILING DATE: 2000-03-03
? NUMBER OF SEQ ID NOS: 114
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 93
? LENGTH: 487
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-808-602-93

```

Query Match	13.1%	Score 493.5	DB 9	Length 487
Best Local Similarity	27.6%	Pred. No. 1e-29		
Matches 142	Conservative 73	Mismatches 194	Indels 106	Gaps 20

```

QY 195 ELSSPEYPRPYKLUSSCYUSLSIEGSEVLIIDREVSVDVEHNEPTDLYOLKIQDREK 254
Db 51 QLTSPGPEPEYKGOESSTODKAKEGFAVRLVF-QDFDLE--PSQDAGOSVITISFGSD 107
QY 255 HGPPCGK-----TLPHRIE--TKSNVTITVTYDDEGDTGKINH7T5AQCPIYMAP 306
Db 108 PSQCGGQGOGSPLGRPPQOREVSSGRELRLFKTPQSEKTAHLH----- 153
QY 307 PNGHVSFVQAKYIILKDSFISFCETGEYELLOGLHPLKSF7AVCOKDGSMDRPMACSI 366
Db 154 -----KGFALITYTAV-----NYSQPISEA----- 174
QY 367 GPDDLESGRVEYITIGCVTTTYKAVIOYSCSEETFYTKVNDGKYVCEADGFWTSSKGEK 426
Db 175 -----SRGEALINAGDNPAAK--YQNHQOEPRY-QAAAGALTCATPCTWKDRDGE 224
QY 427 LPVCEPVC-----LSARTTGGRTYGGQAKKRGDPFWQVYLIGGTTAAGALLDXMVL 480
Db 225 VLOQMPVCGREYTPPIAONQTT---LSSRAKLGNFPMQAFSTIGHGGGALLGDRMILT 280
QY 481 AAHVAVYDCK-----HDAALDIIRMG-----PLKRLSPITYQAMSEAVFIHGGYHDA--GF 529
Db 281 AAHVITYRKDSYSLKKNOSVAVFIQHTAIDELMLKLGHPV-----HRVAVHPDYKQNESHN 336
QY 530 DNDILKILNNKVVINSNTIPICLPKKEAESFMKRTDIDIGTASGMLTQRCFLARNLMYVD 589

```

Db 337 SCGIALLELOHSIFLGNVLPVCLP--DNETLYRSGLLGIYSGFGM-EMGNLTTELKYSR 393
QY 590 IPVDHOKCTAAAYEKPPYPRGSVTANMLCAGLESGKDCSGDSCGALVFLDSETERFEV 649
Db 394 LPVAPREACNAMLQKROPE-VFSDNMFVCYDEQIRHSVCGDGSYYVVDNHAHHVA 452
QY 650 GGIVSWGSMNGEAGGYCVYTKVINYIPMIENIIS 684
Db 453 TGIYSWG-IGCGEG--YDEYTKVLVYDWMIKGVNN 484

RESULT 15

US-09-808-602-21
; Sequence 21, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Heriman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: Macdougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808, 602
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-21

Query Match 13.1% Score 492; DB 9; Length 452;
Best Local Similarity 27.4%; Pred. No. 1,2e-29;

Matches 14; Conservative 74; Mismatches 194; Indels 106; Gaps 20;

QY 195 ELSPYPRPKLSCTYSISLEGFVSVIDFVESPDVETHPETLCYDPLKIQTDREE 254
16 QLTSPGPEPRYKQGESSTDIKAREGFAVLVF-QDFDLE--PSQDCAGDSVTTISFVGS 72
255 HGPFCKG-----TLPHRIE--TKSNVTITFVDESQDHTGKIHVYTAQPCPYMAP 306
Db 73 PSQFCGQGSPLGRPPGQREFVSSGRSLRLTFRQPSSENKTAHLH----- 118
QY 307 PNGHVSVOAKYIILKDSFSICETGYELLOGHLPLKSTAVACQDGSMDRMPACSIYDC 366
Db 119 -----KGFLLYQTVAV-----NYSQPISEA----- 139
QY 367 GPPDLPGRVEYITGPGVTYKAVIOYSCETFEYTKMKVNDGVCEADGWTSSKGEKS 426
Db 140 -----SRGEALINAPGDNDNAK--VQNHCOEPIY-QAAAGALTCATPGTKRQDDEE 189
QY 427 LPVCEPVCG-----LSARTTGRIYGGQAKKPGDFPMQVLLIGTTAAGALLVDMNVL 480
Db 190 VLOGMPVCGRPVTPIAQNOTT-----LSSRAKLGNFPMQATFSTHGRGALLDDBRWILT 245
QY 481 AAHAVYEOK-----HDASALDIRMG-----TLKRLSPHYTQWSEAVFHEGYTHDA--GF 529
Db 246 AARHTYKRDVSLSKKNOSVNVFELGHTAIDEMKLGNHPV---HRVYVHPDYRONESHNF 301
QY 530 DNDIALIKLNKVVYINSNITPICLPKREAESFMRTDDITGSGWGLTORGLFANRLMYVD 589
Db 302 SCGIALLELOHSIFLGNVLPVCLP--DNETLYRSGLLGIYSGFGM-EMGNLTTELKYSR 358
QY 590 IPIVDHOKCTAAAYEKPPYPRGSVTANMLCAGLESGKDCSGDSCGALVFLDSETERFEV 649

Db 359 LPVAPREACNAMLQKROPE-VFSDNMFVCYDEQIRHSVCGDGSYYVVDNHAHHVA 417
QY 650 GGIVSWGSMNGEAGGYCVYTKVINYIPMIENIIS 684
Db 418 TGIYSWG-IGCGEG--YDEYTKVLVYDWMIKGVNN 449

RESULT 16

US-09-808-602-19
; Sequence 19, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Heriman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: Macdougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808, 602
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-19

Query Match 13.1% Score 492; DB 9; Length 487;
Best Local Similarity 27.1%; Pred. No. 1,3e-29;

Matches 144; Conservative 76; Mismatches 206; Indels 106; Gaps 20;

QY 178 RTSGALCSGOYFORSCELSPEYPRPKLSCTYSISLEGFVSVIDFVESPDVETHP 237
Db 34 QACPTRGSVLLAQOLPQQLTSPGPEPRYKQGESSTDIKAREGFAVLVF-QDFDLE--P 90
QY 238 ETLCPYDPLKIQTDREEHGPFCKG-----TLPHRIE--TKSNVTITFVDESQDHTGW 289
Db 91 SQDCAGDSVTTISFVGSDFSCGQGSPLGRPPGQREFVSSGRSLRLTFRQPSSENKTA 150
QY 290 KIHVYTAQPCPYMAPPNGHVSVOAKYIILKDSFSICETGYELLOGHLPLKSTAVACQ 349
Db 151 HLH-----KGFLLYQTVAV-- 165
QY 350 KDGSMDRMPACSIYDCGPPDLPGRVEYITGPGVTYKAVIOYSCETFEYTKMKVNDGK 409
Db 166 ---NYSQPISEA-----SRGEALINAPGDNDNAK--VQNHCOEPIY-QAAAGA 207
QY 410 YVCEADGWTSSKGEKSLPVCEPVCG-----LSARTTGRIYGGQAKKPGDFPMQVLL 463
Db 208 LTCATPGTKRQDDEEYLOGMPVCGRPVTPIAQNOTT-----LSSRAKLGNFPMQATF 263
QY 464 GGTTAAGALLVDMNVLTPAAHAVYEOK-----HDASALDIRMG-----TLKRLSPHYTQW 514
Db 264 IHGRGAGALLDDBRWILTAAHTYKRDVSLSKKNOSVNVFELGHTAIDEMKLGNHPV--- 319
QY 515 EAVFIHEGYTHDA--GFNDIALIKLNKVVYINSNITPICLPKREAESFMRTDDITGASG 572
Db 320 HRVYVHPDYRONESHNFSGGIALLELOHSIFLGNVLPVCLP--DNETLYRSGLLGIYSG 377
QY 573 WGLTORGLFANRLMYVDIPIVDHOKCTAAAYEKPPYPRGSVTANMLCAGLESGKDCSGD 632
Db 378 FGM-EMGNLTTELKYSRLPVAPREACNAMLQKROPE-VFSDNMFVCYDEQIRHSVCGD 435

Db 783 QITPRMAGVFLSGVDSOCDSGGLSSVEAD-GRIFAGVYVWMD-GCAQHNKPGVYT 840
QY 671 KVINIPIWIE 680
Db 841 RLPLFRDWIK 850

RESULT 19
US-09-900-751-2
; Sequence 2, Application US/09900751
; Patent No. US200202653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900,751
; PARENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PR1
; ORGANISM: Mus musculus
US-09-900-751-2

Query Match 12.5%; Score 469.5; DB 10; Length 855;
Best Local Similarity 25.4%; Pred. No. 1.3e-27;
Matches 187; Conservative 101; Mismatches 306; Indels 143; Gaps 33;

QY 9 LIGGSVATPLGPKWPE-----PYFGRLASGPGPE-YANDQERRWLTAP 52
Db 192 VLTSSVAFPIPDPMLORTQDNCSFALHAGAAVTRFTTGFPNSPPAHARQWLRCD 251
QY 53 PGYRLRLFTFHDLELSHCEYDFVRLSSGAK-----VIATLGOESTDTERAPKD-TP 106
Db 252 ADSVLSLTFERSFVAPCDSEHSDLVTVYDLSLSPMHPAAVRLCGTFS-----PSYNLTF 305
QY 107 YSLGSSLDITFRSDYSENEKFTGFEA-FYAABIDECQVAPGAPFCDHCHHNLGFTYC 165
Db 306 LSSQNVFLVLTITNDRHP--GEATPFQOLPKMSSC-----GGF-- 343
QY 166 SCRAGYVLHNRKRTCSALCSGQVFTORSGELSSPEYPRPYPKLSCTYSISLEEGFSYL 225
Db 344 -----LSDTQGFSSPYFGHPPINACTWINKVNNRNKVV 360
QY 226 DEVESDVEETH-PELCPYDFELIQTDREHGPCGKTLPHRIETKSNVTITFTVDESG 284
Db 381 RFLFLYLPVNPVYSGCTKDYVEINSEK-----YCGERSQFVSSNSKLTTFHPSHSXY 435
QY 285 DHGKMIHTS--TAQPCPYMAPRNGHVSPLYAK-----YIKDSFSIFCEGYE 333
Db 436 TDGFLAEVLYSDNDPCPGEMFKTGRCIRKRLCDGMADCPDY--SDERYCRNANTHQ 493
QY 334 LLOGHLPLKSFYAVCOK-----DGSMDR--PMPA-----CSIVDCGPPDDLPSGRVEXITG 382
Db 494 FTCKNGFCRPLFVYVDCSVDNCGSGDEGSCPAAGSFKCNKGLQSOQCNKNDKNGDG 553
QY 383 PGVTTYKAVIYQSCSEETFTYMKVNDKYYC-----EADGEWTSKSG--EKSLLPVECPVC 434
Db 554 SDEASCDVAVVWSC--TKYTYRCONG--LCLSKGNPECDKDTCSGSDSKN--CD--C 604
QY 435 GLSARTTGRIYGGOKAKPDDEPMOVL--LG-CITTAAGLLLDNMVLTAAHAYVEQKH- 490
Db 605 GLKSFTRQARVVGTVNADGEMWQVSLHALGGHLCGASLISPDMLVSAAHCFQDDKNF 664

QY 491 ---DASALDIRMGTL---KRLSPHYTOAMSEAVFIEGTYHDAGFNDIALIKLNKKVYI 544
Db 665 KYSIDYTMWTAFLGLLDQSKRSASGVQELKRIITHPSF--NDTFPYDIALLELESVEY 723
QY 545 NSNITPICLPKREAESFMRTDITGASGWLQRGFL-ARNLMYVDIPIYDHOKCTAAVE 603
Db 724 STVVRPICLP--DATHVFPAGKAIWVGWCHTKEGGTALILQKGEIRVINTTCEDLMP 781
QY 604 KPPYPRGSVTANMLCAGLESGCKDSRGDSGGLVFLFDSTEEWPFVGYISKMSMNGEA 663
Db 782 Q-----QITPRMAGVFLSGVDSOCDSGGLSSAEND-GRMFOAGVWSGE-GCAQR 833
QY 664 GQYGVYTKVINIPIWIE 680
Db 834 NKEGVYTRLPVYVDWIK 850

RESULT 20
US-09-981-123-2
; Sequence 2, Application US/09981123
; Patent No. US20020151029A1
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Human Serine Protease
; FILE REFERENCE: 99-88C1
; CURRENT APPLICATION NUMBER: US/09/981,123
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/167,038
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/715,994
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 269
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-981-123-2

Query Match 11.9%; Score 446.5; DB 10; Length 269;
Best Local Similarity 37.6%; Pred. No. 1.6e-26;
Matches 100; Conservative 41; Mismatches 92; Indels 33; Gaps 8;

QY 444 RITGGKAKRGPDEPMQVLLIGGTTA-----AGALLYDNMVLTAHAHAYVEQKHDA 494
Db 1 RITGGRAEPGLFPWQALIVEDTISRYPNDKMFGSGALLSASWILTRAHILRSQRDTTV 60
QY 495 LDI-RMGTLRLSPHYTOAMSEA-----VFTHEGTYHDAGFNDIALIKLNKKVINS 546
Db 61 IYPSKEHYVLYLGLHDYRDKSGAVNNSAARVYLHPDE-NIQNNYHDLALYQLEPPYLP 119
QY 547 NITPICLPKREAESFMRTDITGASGWLQRGF-----LARNLMYVDIPIYD 594
Db 120 HVMVCLPLREPG-PAPHMLGLVAGWISNPNVTYDEIISGTRFLSDVLYVVKLPVVP 178
QY 595 HOKCTAAVEKPPYPRGSVTANMLCAGLESGCKDSRGDSGGLVFLFDSTEEWPFVGYIS 654
Db 179 HAECGTYESRS--GNYSVENMFCAGYEGGKDTCLSDSGAFVIFDDLSQRWVAGLV 237
QY 655 WGS-MNGEAGQYGVYTKVINIPIW 679
Db 238 WGPPEEGSGSQYGVYTKVSNYVDW 263

RESULT 21
US-09-978-295A-169
; Sequence 169, Application US/09978295A
; Patent No. US2002015606A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan
 APPLICANT: Ferrara, Napoleon
 APPLICANT: Flvaroff, Ellen
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerltsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, J. Christopher
 APPLICANT: Gurney, Austin L.
 APPLICANT: Hillan, Kenneth J.
 APPLICANT: Kijavin, Ivar J.
 APPLICANT: Kuo, Sophia S.
 APPLICANT: Napier, Mary A.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Shelton, David L.
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William I.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE REFERENCE: P2630P1C11
 CURRENT APPLICATION NUMBER: US/09/978,295A
 PRIOR FILING DATE: 2001-10-15
 PRIOR APPLICATION NUMBER: 09/918585
 PRIOR FILING DATE: 2001-07-30
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064249
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/065311
 PRIOR FILING DATE: 1997-11-13
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/077450
 PRIOR FILING DATE: 1998-03-10
 PRIOR APPLICATION NUMBER: 60/077632
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077641
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077649
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077791
 PRIOR FILING DATE: 1998-03-12
 PRIOR APPLICATION NUMBER: 60/078004
 PRIOR FILING DATE: 1998-03-13
 PRIOR APPLICATION NUMBER: 60/078886
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078936
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078939
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/079294
 PRIOR FILING DATE: 1998-03-25
 PRIOR APPLICATION NUMBER: 60/079656
 PRIOR FILING DATE: 1998-03-26
 PRIOR APPLICATION NUMBER: 60/079664
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079689
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079663
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079728
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079786
 PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079920
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 60/079923
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 60/080105
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080107
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080165
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080194
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080327
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080328
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080333
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080334
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/081070
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081049
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081071
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081195
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081203
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081229
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081955
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081817
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081819
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081952
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081838
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/082568
 PRIOR FILING DATE: 1998-04-21
 PRIOR APPLICATION NUMBER: 60/082569
 PRIOR FILING DATE: 1998-04-21
 PRIOR APPLICATION NUMBER: 60/082704
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082804
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082700
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082797
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082796
 PRIOR FILING DATE: 1998-04-23
 PRIOR APPLICATION NUMBER: 60/083336
 PRIOR FILING DATE: 1998-04-27
 PRIOR APPLICATION NUMBER: 60/083322
 PRIOR FILING DATE: 1998-04-28
 PRIOR APPLICATION NUMBER: 60/083392
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083495
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083496
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083499
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083545
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083554
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083558

1 PRIOR FILING DATE: 1998-04-29
 2 PRIOR APPLICATION NUMBER: 60/083559
 3 PRIOR FILING DATE: 1998-04-29
 4 PRIOR APPLICATION NUMBER: 60/083500
 5 PRIOR FILING DATE: 1998-04-29
 6 PRIOR APPLICATION NUMBER: 60/083742
 7 PRIOR FILING DATE: 1998-04-30
 8 PRIOR APPLICATION NUMBER: 60/084366
 9 PRIOR FILING DATE: 1998-05-05
 10 PRIOR APPLICATION NUMBER: 60/084414
 11 PRIOR FILING DATE: 1998-05-06
 12 PRIOR APPLICATION NUMBER: 60/084441
 13 PRIOR FILING DATE: 1998-05-06
 14 PRIOR APPLICATION NUMBER: 60/084637
 15 PRIOR FILING DATE: 1998-05-07
 16 PRIOR APPLICATION NUMBER: 60/084639
 17 PRIOR FILING DATE: 1998-05-07
 18 PRIOR APPLICATION NUMBER: 60/084640
 19 PRIOR FILING DATE: 1998-05-07
 20 PRIOR APPLICATION NUMBER: 60/084598
 21 PRIOR FILING DATE: 1998-05-07
 22 PRIOR APPLICATION NUMBER: 60/084600
 23 PRIOR FILING DATE: 1998-05-07
 24 PRIOR APPLICATION NUMBER: 60/084627
 25 PRIOR FILING DATE: 1998-05-07
 26 PRIOR APPLICATION NUMBER: 60/084643
 27 PRIOR FILING DATE: 1998-05-07
 28 PRIOR APPLICATION NUMBER: 60/085339
 29 PRIOR FILING DATE: 1998-05-13
 30 PRIOR APPLICATION NUMBER: 60/085338
 31 PRIOR FILING DATE: 1998-05-13
 32 PRIOR APPLICATION NUMBER: 60/085323
 33 PRIOR FILING DATE: 1998-05-13
 34 PRIOR APPLICATION NUMBER: 60/085582
 35 PRIOR FILING DATE: 1998-05-15
 36 PRIOR APPLICATION NUMBER: 60/085700
 37 PRIOR FILING DATE: 1998-05-15
 38 PRIOR APPLICATION NUMBER: 60/085689
 39 PRIOR FILING DATE: 1998-05-15
 40 PRIOR APPLICATION NUMBER: 60/085579
 41 PRIOR FILING DATE: 1998-05-15
 42 PRIOR APPLICATION NUMBER: 60/085580
 43 PRIOR FILING DATE: 1998-05-15
 44 PRIOR APPLICATION NUMBER: 60/085573
 45 PRIOR FILING DATE: 1998-05-15
 46 PRIOR APPLICATION NUMBER: 60/085704
 47 PRIOR FILING DATE: 1998-05-15
 48 PRIOR APPLICATION NUMBER: 60/085697

Query Match	11.8%	Score 444:	DB 9:	Length 802:
Best Local Similarity	29.4%:	Pred. No. 1e-25:		
Matches 153:	Conservative 81:	Mismatches 200:	Indels 86:	Gaps 31:
OY	194	GELSSPEYPRPYRKLSSTCTYSI	---SLEEGFVILDFV---	ESFDVETHPETLCSPDFL 246
Db	340	GLVLSPPYPPYSPYSPQTHCSMHLTVLSFLYGLALMDAVALRKQYDLP	-----CTGGGW	393
OY	247	KIQDTREHGPRCG	---KTLPHRETQSNM---	VITLTPYDSEGHDTGMIKH--TSTAP 299
Db	394	TIQNR	-----LCGLRIQLQPYAKERLPVVAITAGITLINFISQLSTLGGVAVNHGLYNQSDP	448
OY	300	CPYP-MAPPNHVSFVQAKYLIKDSFISFCETGYELLOCHLPKLSFTAVQ	-----KDG	353
Db	449	CPGELICSVNGICVP	-ACDGVKD-----CPNG-----	LDERNCRCRATFOCKERS 492
OY	354	WDRPACSIYDCGPRDPLPSGRVYETITGPGVTYTKAVLYQSCSEFTYMKYN	-----DGK	410
Db	493	TCISLP	---KVCD-GQPD	LNSDEGQOEGRPC---GFTTFPCEDRSCYKKNPQCDGAP 546
OY	411	VCLEADGFMSTSGKEKSLPVCPEVGLSARTTGRIYGCOKAKPDP	PWQ-VLLIGTTA	468
Db	547	DCR-DG	---SDEH---CD---CGIQGPS	---RIVGVAGVSEEGMPPWASIQVYGRHIC 593

[illegible]

```

1  RESULT 22
2  US-09-978-697-169
3  : Sequence 169, Application US/09978697
4  Patent No. US20020169284A1
5  GENERAL INFORMATION:
6  APPLICANT: Ashkenazi, Avi
7  APPLICANT: Baker Kevin P.
8  APPLICANT: Bolstein, David
9  APPLICANT: Desnoyers, Luc
10 APPLICANT: Eaton, Dan
11 APPLICANT: Ferrara, Napoleon
12 APPLICANT: Filvaroff, Ellen
13 APPLICANT: Fong, Sherman
14 APPLICANT: Gao, Wei-Qiang
15 APPLICANT: Gerber, Hanspeter
16 APPLICANT: Gerritsen, Mary E.
17 APPLICANT: Goddard, Audrey
18 APPLICANT: Godowski, Paul J.
19 APPLICANT: Grimaldi, J. Christopher
20 APPLICANT: Gurney, Austin L.
21 APPLICANT: Hillan, Kenneth J
22 APPLICANT: Kijavini, Ivar J.
23 APPLICANT: Kuo, Sophia S.
24 APPLICANT: Napier, Mary A.
25 APPLICANT: Pan, James;
26 APPLICANT: Paoni, Nicholas F.
27 APPLICANT: Roy, Margaret Ann
28 APPLICANT: Shelton, David L.
29 APPLICANT: Stewart, Timothy A.
30 APPLICANT: Tumas, Daniel
31 APPLICANT: Williams, P. Mickey
32 APPLICANT: Wood, William T.
33 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
34 FILE REFERENCE: P2630PIC27
35 CURRENT APPLICATION NUMBER: US/09/978,697
36 CURRENT FILING DATE: 2001-10-16
37 PRIOR APPLICATION NUMBER: 09/918585
38 PRIOR FILING DATE: 2001-07-30
39 PRIOR APPLICATION NUMBER: 60/062250
40 PRIOR FILING DATE: 1997-10-17
41 PRIOR APPLICATION NUMBER: 60/064249
42 PRIOR FILING DATE: 1997-11-03
43 PRIOR APPLICATION NUMBER: 60/065311
44 PRIOR FILING DATE: 1997-11-13
45 PRIOR APPLICATION NUMBER: 60/066364
46 PRIOR FILING DATE: 1997-11-21
47 PRIOR APPLICATION NUMBER: 60/077450
48 PRIOR FILING DATE: 1998-03-10
49 PRIOR APPLICATION NUMBER: 60/077632
50 PRIOR FILING DATE: 1998-03-11
51 PRIOR APPLICATION NUMBER: 60/077641
52 PRIOR FILING DATE: 1998-03-11
53 PRIOR APPLICATION NUMBER: 60/077649
54 PRIOR FILING DATE: 1998-03-11
55 PRIOR APPLICATION NUMBER: 60/077791

```


APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC63
CURRENT APPLICATION NUMBER: US/09/999,832A
CURRENT FILING DATE: 2001-10-24
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554

1. PRIOR FILING DATE: 1998-04-29
 2. PRIOR APPLICATION NUMBER: 60/083558
 3. PRIOR FILING DATE: 1998-04-29
 4. PRIOR APPLICATION NUMBER: 60/083559
 5. PRIOR FILING DATE: 1998-04-29
 6. PRIOR APPLICATION NUMBER: 60/083500
 7. PRIOR FILING DATE: 1998-04-29
 8. PRIOR APPLICATION NUMBER: 60/083742
 9. PRIOR FILING DATE: 1998-04-30
 10. PRIOR APPLICATION NUMBER: 60/084366
 11. PRIOR FILING DATE: 1998-05-05
 12. PRIOR APPLICATION NUMBER: 60/084414
 13. PRIOR FILING DATE: 1998-05-06
 14. PRIOR APPLICATION NUMBER: 60/084441
 15. PRIOR FILING DATE: 1998-05-06
 16. PRIOR APPLICATION NUMBER: 60/084637
 17. PRIOR FILING DATE: 1998-05-07
 18. PRIOR APPLICATION NUMBER: 60/084639
 19. PRIOR FILING DATE: 1998-05-07
 20. PRIOR APPLICATION NUMBER: 60/084640
 21. PRIOR FILING DATE: 1998-05-07
 22. PRIOR APPLICATION NUMBER: 60/084598
 23. PRIOR FILING DATE: 1998-05-07
 24. PRIOR APPLICATION NUMBER: 60/084600
 25. PRIOR FILING DATE: 1998-05-07
 26. PRIOR APPLICATION NUMBER: 60/084627
 27. PRIOR FILING DATE: 1998-05-07
 28. PRIOR APPLICATION NUMBER: 60/084643
 29. PRIOR FILING DATE: 1998-05-07
 30. PRIOR APPLICATION NUMBER: 60/085339
 31. PRIOR FILING DATE: 1998-05-13
 32. PRIOR APPLICATION NUMBER: 60/085338
 33. PRIOR FILING DATE: 1998-05-13
 34. PRIOR APPLICATION NUMBER: 60/085323
 35. PRIOR FILING DATE: 1998-05-13
 36. PRIOR APPLICATION NUMBER: 60/085582
 37. PRIOR FILING DATE: 1998-05-15
 38. PRIOR APPLICATION NUMBER: 60/085700
 39. PRIOR FILING DATE: 1998-05-15
 40. PRIOR APPLICATION NUMBER: 60/085689
 41. PRIOR FILING DATE: 1998-05-15
 42. PRIOR APPLICATION NUMBER: 60/085579
 43. PRIOR FILING DATE: 1998-05-15
 44. PRIOR APPLICATION NUMBER: 60/085580
 45. PRIOR FILING DATE: 1998-05-15
 46. PRIOR APPLICATION NUMBER: 60/085573
 47. PRIOR FILING DATE: 1998-05-15
 48. PRIOR APPLICATION NUMBER: 60/085704
 49. PRIOR FILING DATE: 1998-05-15
 50. PRIOR APPLICATION NUMBER: 60/085697

11.8%;	Score 444;	DB 9;	Length 802;
Best Local Similarity 29.4%;	Pred. No. 1e-25;		
Matches 153;	Conservative 81;	Mismatches 200;	Indels 86;
			Gaps 31.

QY	194	GELSPPEYPRPRPKLSCSTYSI---	SLEEGFVILDEV---	ESFDVENHPTCLSPYDEL	246
Db	340	GVLSPPEYPRPSYSPOTHS	SMHLTVSLDGLAMFDAYALRQKIDLP----	CLQGW	393
QY	247	KIQDREHGPRG----	KILPRILETSNT--VITFTYIDESGDHTGKINH--	ISTAOP	299
Db	394	TIQNR-----	LCGRILQPYAERIPVAVATAGITINFTSQISLTG	GGVAVHGLYNQSPR	448
QY	300	CPYP--MAPPNHVSIVQAKYL	IKDSFSLFCERGVELLOGLHPLKSPFAVCO-----	KDSS	353
Db	449	CPGEFLCSVNGLCVP--	ACDGYKD-----	CPRG-----	
QY	354	WDRPMACSIYDCGPPDPLSGRVEY	ITGPGYTVKAVIAYSCEETFTYMKVN-----	DGRY	410
Db	493	TCISLP--KVCYD-GQPPDL-	NSGDEEQOQEGVPC--GTFTFQCEDRSCVYKRNPCQDGRP		546
QY	411	VCADGPFMTSSGKESL	LVCEPVCCLSA	RTTGGRITYGOKAKRGPDPNQ--VLLIGGTTA	468

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 22:02:56 : Search time 37 Seconds

(without alignments)
768,994 Million cell updates/sec

Title: US-09-874-198-2

Perfect score: 3758
Sequence: 1 MRLTLRLGLCGSVATPLGP.....GYTRVINYIPWIENTISDF 686

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 75 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3745	99.7	686	MAS2_HUMAN	O00187 homo sapien
2	1549.5	41.2	704	CRAR_MOUSE	P98064 mus musculu
3	1544	41.1	699	CRAR_HUMAN	P48740 h complimen
4	1412	37.6	705	CIR_HUMAN	P00736 homo sapien
5	1244.5	33.1	695	CASP_MESAU	P15156 mesocricetu
6	1238	32.9	688	CIS_HUMAN	P09871 homo sapien
7	488	13.0	855	ST14_HUMAN	O9Y5Y6 homo sapien
8	471.5	12.5	1019	LFC_TACTR	P28175 tachyples
9	470.5	12.5	1019	LFC_CARRO	Q26422 carcinoscor
10	469.5	12.5	855	ST14_MOUSE	P56677 mus musculu
11	443	11.8	1019	ENTK_HUMAN	P98073 homo sapien
12	422	11.2	475	PA10_CHICK	P25155 gallus gall
13	419	11.1	347	HPT_RAT	P06866 rattus norv
14	418.5	11.1	1034	ENTK_PIG	P98074 sus scrofa
15	416.5	11.1	991	BMPI_MOUSE	P98063 mus musculu
16	412	11.0	452	FA9_CANFA	P19540 canis fami
17	411.5	10.9	986	BMPI_HUMAN	P13497 homo sapien
18	408.5	10.9	1035	ENTK_BOVIN	P98072 bos taurus
19	405	10.8	461	FA9_HUMAN	P00740 homo sapien
20	403	10.8	618	THRB_MOUSE	P19221 mus musculu
21	403	10.7	329	HPT_CANFA	P19006 canis fami
22	403	10.7	488	FA10_HUMAN	P00742 homo sapien
23	402.5	10.7	285	FA9_CAVPO	P16295 cavia porce
24	402	10.7	347	HPT_ATEGE	P50417 atelaf geof
25	401.5	10.7	416	HEPS_MOUSE	O05511 rattus norv
26	400.5	10.7	459	FA9_MOUSE	P16294 mus musculu
27	399.5	10.6	406	HPT2_HUMAN	P00738 homo sapien
28	399.5	10.6	416	FA9_BOVIN	P00741 bos taurus
29	399.5	10.6	416	HEPS_MOUSE	O35453 mus musculu
30	399	10.6	458	PRIC_MOUSE	O28661 oryctolagus
31	397	10.6	461	PRIC_RABIT	P31394 rattus norv
32	395.5	10.5	1069	ENTK_RAT	P97435 mus musculu
33	393.5	10.5	271	FA9_PIG	P16293 sus scrofa

34	393.5	10.5	444	FA7_RABIT	P98139 oryctolagus
35	389.5	10.4	282	FA9_RAT	P16296 rattus norv
36	389	10.4	459	PRIC_PIG	O91422 sus scrofa
37	387	10.3	347	HPT1_HUMAN	P00737 homo sapien
38	387	10.3	625	THRB_BOVIN	P00735 bos taurus
39	385.5	10.3	1022	TLID_BRARE	O57460 brachydantio
40	385	10.2	348	HPT2_HUMAN	P00739 homo sapien
41	384	10.2	347	HPT1_MOUSE	O62558 mus saxicol
42	384	10.2	461	PRIC_MOUSE	P33587 mus musculu
43	383.5	10.2	275	FA9_RABIT	P16292 oryctolagus
44	382	10.2	461	PRIC_HUMAN	P04070 homo sapien
45	382	10.2	622	THRB_HUMAN	P00734 homo sapien
46	381.5	10.2	417	HEPS_HUMAN	P05981 homo sapien
47	376.5	10.0	456	PRIC_BOVIN	P00745 bos taurus
48	376.5	10.0	638	KAL_MOUSE	P26262 mus musculu
49	374	10.0	1057	TLID_DROME	P25723 drosophila
50	369	9.8	492	FA10_BOVIN	P00743 bos taurus
51	368.5	9.8	490	FA10_RABIT	O19045 oryctolagus
52	368.5	9.8	810	PLMN_MACMU	P12545 macaca mula
53	367.5	9.8	407	FA7_BOVIN	P22457 bos taurus
54	367	9.8	259	SAT3_DERFA	P49275 dermatophag
55	367	9.8	638	KAL_HUMAN	P03952 homo sapien
56	366.5	9.8	342	PSS8_MOUSE	O9601 mus musculu
57	366	9.7	261	DER3_DERPT	P39675 dermatophag
58	366	9.7	437	TMS4_HUMAN	O9184 homo sapien
59	362.5	9.6	342	PSS8_RAT	O9687 rattus norv
60	362	9.6	343	PSS8_HUMAN	O1651 homo sapien
61	361	9.6	466	FA7_HUMAN	P08709 homo sapien
62	360.5	9.6	617	THRB_RAT	P18292 rattus norv
63	360.5	9.6	810	PLMN_HUMAN	P00742 homo sapien
64	360	9.6	790	PLMN_PIG	P06867 sus scrofa
65	359	9.6	324	TEST_MOUSE	O9147 mus musculu
66	357	9.5	376	FA10_TROCA	P81428 troglodchis
67	357	9.5	812	PLMN_BOVIN	P06868 bos taurus
68	356.5	9.5	638	KAL_RAT	P14272 rattus norv
69	355.5	9.5	274	FA9_SHEEP	P16291 ovis aries
70	353	9.4	261	EUM3_EURMA	O97370 euryglyphus
71	352.5	9.4	707	BMPI_XENLA	P98070 xenopus lae
72	352	9.4	812	PLMN_MOUSE	P20918 mus musculu
73	351	9.3	375	PCE_TACTR	P21902 tachyples
74	350	9.3	321	TRIG_HUMAN	O9122 homo sapien
75	350	9.3	490	TMS2_MOUSE	O9148 mus musculu

ALIGNMENTS

RESULT 1
MAS2_HUMAN STANDARD: PRT; 686 AA.
AC O00187: 075754: 09Y270: 09B2H0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mannan-binding lectin serine protease 2 precursor (BC 3.4.21.-)
DE (Mannose-binding protein associated serine protease 2) (MASP-2)
DE (MBL-associated serine protease 2).
GN MASP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97242412; PubMed=9087411;
RA Thiel S., Jensen T.V., Stover C.M., Schwaebler W.J., Laurson S.B.,
RA Poulsen K., Willis A.C., Eggleton P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "A second serine protease associated with mannan-binding lectin that
activates complement";
RL Nature 386:506-510(1997).
[2]

RP SEQUENCE FROM N.A.
RA Thiel S., Vorup-Jensen T., Stover C.M., Schwaebble W., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "Identification and characterization of a novel protein of the human
RT complement system, mannan-binding lectin-associated serine protease-2
RT (MASP-2).";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE-99192764; PubMed-10092804;
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,
RA Jensenius J.C., Schwaebble W.J.;
RT "Two constituents of the initiation complex of the mannan-binding
RT lectin activation pathway of complement are encoded by a single
RT structural gene";
RL J. Immunol. 162:3481-3490(1999).
RN [4]
RN SEQUENCE FROM N.A.
RA Park D., Kim B., Baek K., Yoon J.;
RX Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRYPsin PROTEASE THAT PRESUMABLY PLAYS AN IMPORTANT ROLE
CC IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER
CC ACTIVATION IT CLEAVES C4 GENERATING C4a AND C4b.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y09926; CAA71059.1; -
DR EMBL: X98400; CAA67050.1; -
DR EMBL: Y18287; CAB50735.1; -
DR EMBL: Y18286; CAB50733.1; -
DR EMBL: AF321562; AAG50274.1; -
DR EMBL: AF321558; AAG50274.1; JOINED.
DR EMBL: AF321559; AAG50274.1; JOINED.
DR EMBL: AF321560; AAG50274.1; JOINED.
DR EMBL: AF321561; AAG50274.1; JOINED.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.229; -
DR Trimm: HGNC:6902; MASP2.
DR Trimm: 605102; -
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PR00084; sushi_2.
DR Pfam: PR00089; trypsin_1.
DR Pfam: PR00431; CUB_2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP_2.
DR SMART: SM00042; CUB_2.
DR SMART: SM00179; EGF_Ca_1.
DR SMART: SM00020; Tryp_Spec_1.
DR PROSITE: PS00010; ASX_HYDROXYL_1.
DR PROSITE: PS01180; CUB_2.
DR PROSITE: PS01186; EGF_2_1.
DR PROSITE: PS01187; EGF_Ca_1.
DR PROSITE: PS02040; TRYPSIN_DOM_1.
DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE: PS00135; TRYPSIN_SER; 1.

KW Hydrolase; Complement pathway; Serine protease; Protease;
KW Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.
FT SIGNAL 1 15
FT CHAIN 16 686
FT CHAIN 16 444
FT CHAIN 16 444
FT CHAIN 445 686
FT CHAIN 16 137
FT DOMAIN 138 181
FT DOMAIN 184 296
FT DOMAIN 299 362
FT DOMAIN 365 431
FT DOMAIN 445 686
FT ACT_SITE 532 532
FT ACT_SITE 633 633
FT ACT_SITE 158 158
FT MOD_RES 72 90
FT DISULFID 142 156
FT DISULFID 152 165
FT DISULFID 167 180
FT DISULFID 184 211
FT DISULFID 241 259
FT DISULFID 300 348
FT DISULFID 328 361
FT DISULFID 366 412
FT DISULFID 396 430
FT DISULFID 434 552
FT DISULFID 598 618
FT DISULFID 629 660
FT CONFLICT 361 362
FT CONFLICT 371 371
FT CONFLICT 372 372
FT CONFLICT 442 442
FT CONFLICT 447 447
FT CONFLICT 461 462
FT CONFLICT 473 473
SQ SEQUENCE 686 AA; 75685 MW; 4E34DED159448A2A CRC64;
Query Match 99.7%; Score 3745; DB 1; Length 686;
Best Local Similarity 99.7%; Pred. No. 5,2e-274;
Matches 684; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MRLTLTGLIGSGVAPPLGKMPKPEVGRGLASPGFGEYANDDERMTLTAPGRRLRLV 60
Db 1 MRLTLTGLIGSGVAPPLGKMPKPEVGRGLASPGFGEYANDDERMTLTAPGRRLRLV 60
QY 61 FTAFDELISHLCEYDPEVKLSGSAKVLAATLGGESTDTERAPGKDTFYSLGSSLDITFRSD 120
Db 61 FTAFDELISHLCEYDPEVKLSGSAKVLAATLGGESTDTERAPGKDTFYSLGSSLDITFRSD 120
QY 121 YSNEKPPPTGFEATYAAEDIDECOVAAGEAPTCDHCHNHGGRYCSGRAGYVLRNKRKTC 180
Db 121 YSNEKPPPTGFEATYAAEDIDECOVAAGEAPTCDHCHNHGGRYCSGRAGYVLRNKRKTC 180
QY 181 SALCSGQVFTQSGELSSPEYPRPKLSCTYSISLEEGFSVILDPVSEFVDETPELT 240
Db 181 SALCSGQVFTQSGELSSPEYPRPKLSCTYSISLEEGFSVILDPVSEFVDETPELT 240
QY 241 CPYDFLKIOTDREHGFPCGKTLPHRIETKSNVTITFTVDESDDHTGKINHTSTAOPC 300
Db 241 CPYDFLKIOTDREHGFPCGKTLPHRIETKSNVTITFTVDESDDHTGKINHTSTAOPC 300
QY 301 PYMAPPNHVSPOAKYILKDSFISFCEYGLDGHLPKFTAVCOKDGSDMPRA 360
Db 301 PYMAPPNHVSPOAKYILKDSFISFCEYGLDGHLPKFTAVCOKDGSDMPRA 360
QY 361 CSTVDGPPDDLPGRGVEYTGVTYKAVIOYSCSEFTYTKVNDGKYVCADGFWTS 420
Db 361 CSTVDGPPDDLPGRGVEYTGVTYKAVIOYSCSEFTYTKVNDGKYVCADGFWTS 420
QY 421 SKGEKSLPVECPVGLSARTTGRIYGOKAKPGDEPWOVLIGGTAAAGALLYDMWVL 480

```

|||||
Db 421 SKGKSLPVEPCVGLSARTTGGRIYGGQKAKPCDPPMQVLIGCTTAAGALLIDNMVLT 480
Oy 481 AAHAVYQKDKDASLDIRMGITIKRLSPHYQANSEAVFHEGYHDDGPNDAITLTLNN 540
Db 481 AAHAVYQKDKDASLDIRMGITIKRLSPHYQANSEAVFHEGYHDDGPNDAITLTLNN 540
Oy 541 KVINSNITPICLPKREAESFMRTDITGASGMGLTORGFLARNLMTVDIPIVDHQCTA 600
Db 541 KVINSNITPICLPKREAESFMRTDITGASGMGLTORGFLARNLMTVDIPIVDHQCTA 600
Oy 601 AYEKPPRPSVYANMCALESQKDCRGDSGALVFLDSETERMPVGGIVSWGSMNC 660
Db 601 AYEKPPRPSVYANMCALESQKDCRGDSGALVFLDSETERMPVGGIVSWGSMNC 660
Oy 661 GEAGQGVYTKVINYIPMIENITSD 686
Db 661 GEAGQGVYTKVINYIPMIENITSD 686
CRAR_MOUSE
ID CRAR_MOUSE STANDARD: PRT: 704 AA.
AC P98064;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement-activating component of Ra-reactive factor precursor
  (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (RarF)
DE (Mannan-binding lectin serine protease 1).
DE MASPI OR CRARF.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eultheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE=Liver;
RX MEDLINE=94179811; PubMed=8133044;
  Takayama Y., Takada F., Takahashi A., Kawakami M.;
  "A 100-kDa protein in the C4-activating component of Ra-reactive
  factor is a new serine protease having module organization similar to
  C1r and C1s.";
  J. Immunol. 152:2308-2316(1994).
[2]
SEQUENCE OF 465-704 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE=Liver;
RX MEDLINE=93176166; PubMed=8439319;
  Takahashi A., Takayama Y., Hatusue H., Kawakami M.;
  "Presence of a serine protease in the complement-activating component
  of the complement-dependent bactericidal factor, RarF, in mouse
  serum.";
  Biochem. Biophys. Res. Commun. 190:681-687(1993).
[3]
FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF
  WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
  CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
  CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
  C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
[4]
SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
  (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
  CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (29)
  LINKED BY A DISULFIDE BOND.
[5]
TISSUE SPECIFICITY: LIVER.
[6]
DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
[7]
SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
[8]
SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation-
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D16492; BAA03944.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.198; -.
DR MGD: MGI:88492; Maspl.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; Cub_domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF_Ca.
DR InterPro: IPR001881; EGF-like.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi_2.
DR Pfam: PF00089; trypsin_1.
DR Pfam: PF00431; trypsin_2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP_2.
DR SMART: SM00042; CUB_2.
DR SMART: SM00179; EGF_CA_1.
DR SMART: SM00020; TRYP_SPC_1.
DR PROSITE: PS00010; ASX_HYDROXYL_1.
DR PROSITE: PS01180; CUB_2.
DR PROSITE: PS01186; EGF_2_1.
DR PROSITE: PS01187; EGF_CA_1.
DR PROSITE: PS50240; TRYPsin_DOM_1.
DR PROSITE: PS00134; TRYPsin_HIS_1.
DR PROSITE: PS00135; TRYPsin_SER_1.
KW Hydrolyase; Complement pathway; Serine protease;
  Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.
FT SIGNAL 1 24
FT CHAIN 25 704
FT CHAIN 25 453
FT CHAIN 454 704
FT DOMAIN 25 143
FT DOMAIN 144 187
FT DOMAIN 190 302
FT DOMAIN 305 368
FT DOMAIN 371 438
FT DOMAIN 454 704
FT ACT_SITE 495 495
FT ACT_SITE 557 557
FT ACT_SITE 651 651
FT MOD_RES 164 164
FT DISULFID 78 96
FT DISULFID 148 162
FT DISULFID 158 171
FT DISULFID 173 186
FT DISULFID 190 217
FT DISULFID 247 265
FT DISULFID 306 354
FT DISULFID 334 367
FT DISULFID 372 419
FT DISULFID 402 437
FT DISULFID 441 577
FT DISULFID 619 636
FT DISULFID 647 677
FT CARBOHYD 54 54
FT CARBOHYD 183 183
FT CARBOHYD 390 390
FT CARBOHYD 412 412
SQ SEQUENCE 704 AA: 79895 MW: 7144F301D2D2C67F CRC64;
Query Match 41.2%; Score 1549.5; DB 1; Length 704;
Best Local Similarity 42.8%; Pred. No. 6,7e-109;
Matches 295; Conservative 129; Mismatches 212; Indels 53; Gaps 15;
Oy 26 VEGRLASPGPEGYANDQERRMTLTAPPGYRLRLYTHFDLELSHLCYDFVKLSGAKV 85
Db 32 MFCIOQSPGYDSDYSPDSEVMTWNTVPEGFRIKILYFMHFNLESSYLCEYDVVKVETEDQV 91

```


DR EMBL: AB007612; BAA89206.1; JOINED.
DR EMBL: AB007613; BAA89206.1; JOINED.
DR EMBL: AB007614; BAA89206.1; JOINED.
DR EMBL: AB007615; BAA89206.1; JOINED.
DR EMBL: AB007616; BAA89206.1; JOINED.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.198; .
DR Genes: HGNC:6901; MASP1.
DR MIM: 600521; .
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00084; sush1; 2.
DR Pfam: PF00431; CUB; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KM Hydroxylase: Complement pathway: Serine protease: Protease:
KM Glycoprotein: Sush1: Repeat: Signal: EGF-like domain: Hydroxylation.
FT SIGNAL 1 19
FT CHAIN 20 699
FT 20 448
FT CHAIN 449 699
FT DOMAIN 20 138
FT DOMAIN 139 182
FT DOMAIN 185 297
FT DOMAIN 300 363
FT DOMAIN 366 433
FT DOMAIN 449 699
FT ACT_SITE 490 490
FT ACT_SITE 552 552
FT ACT_SITE 646 646
FT MOD_RES 159 159
FT DISULFID 73 91
FT DISULFID 143 157
FT DISULFID 153 166
FT DISULFID 168 181
FT DISULFID 185 212
FT DISULFID 242 260
FT DISULFID 301 349
FT DISULFID 329 362
FT DISULFID 367 414
FT DISULFID 397 432
FT DISULFID 436 572
FT DISULFID 614 631
FT DISULFID 642 672
FT CARBOHYD 49 49
FT CARBOHYD 178 178
FT CARBOHYD 385 385
FT CARBOHYD 407 407
FT CARBOHYD 432 432
FT CARBOHYD 449 449
FT CARBOHYD 499 499
FT CARBOHYD 527 527
FT CARBOHYD 543 543
FT CARBOHYD 552 552
FT CONFLICT 543 543
FT CONFLICT 552 552

FT CONFLICT 643 643 A -> S (IN REF. 1).
SQ SEQUENCE 699 AA: 79258 MW: ADD9697AE6AB01B5 CRC64;
Query Match 41.18: Score 1544; DB 1: Length 699;
Best Local Similarity 42.28: Pred. No. 1.7e-108;
Matches 300; Conservative 131; Mismatches 234; Indels 46; Gaps 14;
QY 1 MRLTLGLGCGSVATPLGPKWP-EPVFGRGLASGPGGEVANDERRHMTLPAPGYRL 59
DB 1 MRWLLLYALCFSLSKSANTVELNNHFGQIQSGYGPDSYSDSEVYWNITVPDGFRL 60
QY 60 YTFHFDLSHLCEYDEVKLSGAKVLATLGGQSTPTTERAPCKDPTFSLSSDIIFRS 119
DB 61 YFMHFNLESSYLCEYDYKVETEDQVLATFCGRRTDTEDQPGGEVYLSPEFSMIFRS 120
QY 120 DYSNEKPTGFEAFYAAEDIDECOVAPGEAPTCOHCHNHILGFGYCSGRACYLHRRKRT 179
DB 121 DFNSEERFTGDAHYMAVDDECKEREDEELSCDHCHNYIGYGYCSRFEGYILHPTDNR 180
QY 180 CSAICSGOVFORSGELSSPEYPRPKLSCTSYSLSEGEFVILDPVESFDVETPET 239
DB 181 CRVCSQNLFTQRTGVTSPDPENPYRPSSECLTYILEBGFVNLQPEDIFDIQDHEV 240
QY 240 LCPYDFLKIQTDRREHGFPGCKTLPHRIETKSNFTVTTFVYDESGDHTGKIHVYSTAOP 299
DB 241 PCPYDYIKIKKQPRVLPFGCEKARPEISTOSHVLILFHSDNAENKGRILSYRAAGNE 300
QY 300 CPYPMAPNGHVSVPVQAKYILKDSFISFCETGYLLQGHLPKSFYAVCQKGSMDRPM 359
DB 301 CPRLQPRVHGKIEPSQAKYFFKQDVLVSCDGYKVLKDNVEMDFQIECLKDGTMSKIP 360
QY 360 ACSIVDGPDPDPSGVEYITGCVTTKAVITQYSCSEETFTYK-KVNDGKYCEADGFW 418
DB 361 TCKIVDCRAGELHSLTSTSTRNNLTYYSEIKYSCDEPKLNNNTGLITYSAGVW 420
QY 419 TSSKGEKSLPCEPVCGLS--ARTTGRIVYGQAKRQDPFQVLLI--GTTAAGAL 473
DB 421 MNKVLGSLPLFCVCGLPKRSKRLMARIFNGRAQAGTTPWIMLSLNLQPPCGSGL 480
QY 474 YDNVNLTAANVYQKHDASLDTRMGTLK--RLSP-----HTQAMSEAVFHGEG 522
DB 481 GSSIVTAHACLHQ-----SLDKPDKPLRSDLSPSDFKILIGKHRLRSLDEHQL 534
QY 523 YTHA-----GDDNDIALIKNNKRVINSNTPICLPKKAESEFRTDICTAGSG 574
DB 535 VKHTTLHPQIDPNTFEDNALVELLESPLVAFWPCLP---EGPOEGAMVYVSGW 590
QY 575 LTQGFPLAR--NMYVDIPVDHQKCTAAVEKPPYRGSVYANMLCAGLESGKDSGRC 631
DB 591 ---KQFLORFETLMEIEIPVDHSTOCKAY--APLAK-KVTRMICAGEGKDKACAG 644
QY 632 DSGCALVFLDSETERMFVGVIVSGMNGCEAGGYVYTKVINYIPENT 662
DB 645 DSGCPMVTLNREGQVYLVGVSMGD--DCGKRDRYGVSYIHNKHWIORV 694
RESULT 4
CIR_HUMAN STANDARD; PRT; 705 AA.
ID AC P00736;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement C1r component precursor (EC 3.4.21.41).
GN C1R.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87026566; Pubmed=3021205;
RA Leytus S.P., Kurachi K., Sakarlassen K.S., Davie E.W.;

RT "Nucleotide sequence of the cDNA coding for human complement C1r.";
 RL Biochemistry 25:4855-4863(1986).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87156625; Pubmed=3030286;
 RA Journet A., Tost M.;
 RT "Cloning and sequencing of full-length cDNA encoding the precursor of
 RL human complement component C1r.";
 RL Biochem. J. 240:783-787(1986).
 RN [13]
 RP SEQUENCE OF 18-463.
 RX MEDLINE=87241248; Pubmed=3036070;
 RA Atlaud G.J., Willis A.C., Gagnon J.;
 RT "Complete amino acid sequence of the A chain of human complement-
 RL classical-pathway enzyme C1r.";
 RL Biochem. J. 241:711-720(1987).
 RN [14]
 RP SEQUENCE OF 464-705.
 RX MEDLINE=83204782; Pubmed=6303394;
 RA Atlaud G.J., Gagnon J.;
 RT "Complete amino acid sequence of the catalytic chain of human
 RL complement subcomponent C1-r.";
 RL Biochemistry 22:1758-1764(1983).
 RN [15]
 RP SEQUENCE OF 152-186, AND HYDROXYLATION.
 RX MEDLINE=88005128; Pubmed=2820791;
 RA Atlaud G.J., van Dorselaer A., Bell A., Mancini M., Aude C.,
 RA Gagnon J.;
 RT "Identification of erythro-beta-hydroxyasparagine in the EGF-like
 RL domain of human C1r.";
 RL FEBS Lett. 222:129-134(1987).
 RN [16]
 RP SEQUENCE OF 133-137; 187-211 AND 609-613, AND PHOSPHORYLATION.
 RX MEDLINE=96221263; Pubmed=8635594;
 RA Pelloux S., Thieles N.M., Hudry-Clergeon G., Petitlot Y., Filhol O.,
 RA Atlaud G.J.;
 RT "Identification of a cryptic protein kinase CK2 phosphorylation site
 RL in human complement protease C1r, and its use to probe intramolecular
 RT interaction.";
 RL FEBS Lett. 386:15-20(1996).
 RN [17]
 RP STRUCTURE BY NMR OF 140-192.
 RX MEDLINE=9818432; Pubmed=9477945;
 RA Bersch B., Hernandez J.-F., Marion D., Atlaud G.J.;
 RT "Solution structure of the epidermal growth factor (EGF)-like module
 RL of human complement protease C1r, an atypical member of the EGF
 RT family.";
 RL Biochemistry 37:1204-1214(1998).
 RL -1- FUNCTION: C1R B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1O
 AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
 OF THE COMPLEMENT SYSTEM.
 -1- CATALYTIC ACTIVITY: ACTIVATES C2 BY PROTEOLYTIC CLEAVAGE SO
 THAT IT CAN, IN TURN, ACTIVATE C3 AND C4.
 -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1O, R
 AND S IN THE MOLAR RATION OF 1:2:2. C1R IS A DIMER OF IDENTICAL
 CHAINS, EACH OF WHICH IS ACTIVATED BY CLEAVAGE INTO TWO CHAINS, A
 AND B, CONNECTED BY DISULFIDE BONDS.
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04701; CAA28407.1; -;
 DR EMBL: M14056; AAA51851.1; -;
 DR PIR: A24170; C1H0RB.

DR PDB: 1APQ; 17-SEP-97.
 DR MEROPS: S01.192; -;
 DR Genew: HGNC:1246; C1R.
 DR MIM: 216950; -;
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; Tryp_Spc; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Complement pathway; Plasma; Glycoprotein; Serine protease; Hydrolase;
 KW Hydroxylation; Phosphorylation; Sushi; Repeat; Signal;
 KW EGF-like domain; 3D-structure.
 KW SIGNAL
 FT CHAIN 18 463 COMPLEMENT C1R HEAVY CHAIN.
 FT CHAIN 464 705 COMPLEMENT C1R LIGHT CHAIN.
 FT DOMAIN 18 141 CUB 1.
 FT DOMAIN 142 190 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 193 305 SUSHI 1.
 FT DOMAIN 308 372 SUSHI 2.
 FT DOMAIN 375 448 SERINE PROTEASE.
 FT DOMAIN 464 705 CHARGE RELAY SYSTEM.
 FT ACT_SITE 502 502 CHARGE RELAY SYSTEM.
 FT ACT_SITE 557 557 CHARGE RELAY SYSTEM.
 FT ACT_SITE 557 557 CHARGE RELAY SYSTEM.
 FT ACT_SITE 654 654 CHARGE RELAY SYSTEM.
 FT MOD_RES 167 167 HYDROXYLATION.
 FT MOD_RES 206 206 PHOSPHORYLATION (BY CK2).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .).
 FT DISULFID 71 89 PROBABLE.
 FT DISULFID 146 165 PROBABLE.
 FT DISULFID 161 174 PROBABLE.
 FT DISULFID 176 189 PROBABLE.
 FT DISULFID 193 220 PROBABLE.
 FT DISULFID 250 268 PROBABLE.
 FT DISULFID 309 358 PROBABLE.
 FT DISULFID 338 371 PROBABLE.
 FT DISULFID 376 429 PROBABLE.
 FT DISULFID 406 447 PROBABLE.
 FT DISULFID 451 577 INTERCHAIN (PROBABLE).
 FT DISULFID 620 639 POTENTIAL.
 FT DISULFID 680 680 POTENTIAL.
 FT CONFIDET 152 152 S -> L (IN REF. 2).
 SQ SEQUENCE 705 AA; 5CBCC0201061463 CRC64;
 Query Match 37.6%; Score 1412; DB 1; Length 705;
 Best Local Similarity 41.7%; Pred. No. 1,5e-98;
 Matches 301; Conservative 108; Mismatches 248; Indels 64; Gaps 22;
 QY 1 MRLTLTL-GLLC---GSVATPLGPKWPEYFGRGLASPGPGYANDQERRMTLTAPPGY 55
 DB 1 MWLLTLVLPALFCRAGSI--PI---PQKLFGEVSPLEPKPYPNNEFTTVIVPTGT 54
 QY 56 RLRLYFTHEPDELSEHUCEYDFVFLSSGAKVLAFLCGQESTDTERADPGKDTFSLGSSLDI 115


```

116 TERSDYENK-----FTGFAFAAADIDEC--QVAPGE--APICDHCHNHLGFEYC 165
115 TFRHDSNEEGTMTFKGLAYQAADLDCASRSKSGEEDPOPCQCHLCHNVGVGYFC 174
166 SCRAGYVLHNRKRTCSALCSGOVFTQSGELSPERYPRPKLSYCSYLSLEGFEVYL 225
175 SCRRGYELQEDRHSCQACSESELTEASGYSSLEYRSPYRDLRCYNSIVEGGLLHL 234
226 DPESDVEVETHPETLCYDFELKIQTDREHGPCCKTLPHRIETKSNVTYTFVDES 285
235 KLEPFPIDHQQVHCYDLOLQIYANCKNIGFEFCGKORPDLDTSSNAVDLFTDES 294
286 HTGKHYTSAOCPRYPMAPN--GHVSPVQAKYILKDSFISFETCYELLOGLPLKSF 344
295 SRGKMLTYTEILKCPKPKLIDFTTIIQNLQPOYQFNDYFATCKQGLLEGNOVLHS 354
345 TAVCQKDSMDRPMRACSIYDCGPPDLPSGRVEYITGPGVYTKAVIYSCSETFYTM 404
355 TAVCGDDGTMRAMPKCKINDCCGPNLNDPFRYTTMGVNTYKARIQYCHNEYKMQ 414
405 V-----NDGKYVCEADGEW-TSSKGEKSLPCEPVCG--LSARTTGGRIYGOQAK 454
415 TRASRESEGVYTCAGIWKKNKQKGEK-IPRCLPVCGKVPNVEDRORIIGQKAKMG 473
455 DEPMQVILGTTAGALLDNVLTAAHAYBEQKHA---SALDIMG-----TKRLS 506
474 NEPMQVNTNHGRGGALLGGRMLTAHLYTPKEHEQSNASIDVGLGHTNVEELKLG 533
507 PHYQANSEAVFIHEGYTHDA--GFNDIALIKLNKVINSNTPICLPKEASEFMT 564
534 NHPIR---RVSVPDYRDESYNFEEDIALLELNSVTGLGNLPLICLP--DNDTFYDL 587
565 DDITAGKGLTQNGFLARNLMTYDIPYDHQKCTAAIEKPYRGS-----VTANMLCA 619
588 GLMVGVGFGVMEER-IAHDLRFVRLPVANPOAC-----ENMLRGRNMDVFSQNFCA 640
620 GLEGGKDCSGDSCGALVFLDSETRMFPVGGIYSGMSMNGEAGOGVYVKNYIPWI 679
641 GHPLKDCGDSGSGVFAVRDPMTDRMVAITGIVSWG-IGCSRG--YGFYTKVLTNVDWI 697
QY 680 E 680
Db 698 K 698
T 5
MESAU STANDARD: PRT: 695 AA.
CASP_MESAU
P15156:
AC 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calcium-dependent serine proteinase precursor (EC 3.4.21.-) (CASP).
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-50 AND 446-472.
RC TISSUE=Fibroblast; PubMed=2753140;
RX MEDLINE=89325606; PubMed=2753140;
RA Kinoshita H., Sakiyama H., Tokunaga K., Imajo-Ohtani S., Hamada Y.,
RA Isono K., Sakiyama S.;
RT Complete primary structure of calcium-dependent serine proteinase
RT capable of degrading extracellular matrix proteins.*;
RL FEBS Lett. 250:411-415(1989).
CC -1- FUNCTION: CAPABLE OF DEGRADING EXTRACELLULAR MATRIX PROTEINS.
CC CASP DEGRADES TYPE I AND IV COLLAGEN AND FIBRONECTIN IN THE
CC PRESENCE OF CALCIUM.

```

```

CC -1- SUBUNIT: HETERODIMER, CONSISTING OF HEAVY AND LIGHT CHAINS WITH
CC DISULFIDE BONDS. THE HEAVY CHAIN IS EXPECTED TO BE A REGULATORY
CC SUBUNIT AND THE LIGHT CHAIN CONTAINS THE CATALYTIC SITE.
CC -1- DOMAIN: THE GLU-RICH REGION IN THE N-TERMINAL REGION MAY BE GAMMA
CC CARBOXYLATED AND FUNCTION AS A CALCIUM-BINDING SITE.
CC -1- SIMILARITY: TO BLOOD COAGULATION FACTORS SUCH AS IX, X AND AN
CC ANTICOAGULATION FACTOR, PROTEIN C.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16160; CAA34286.1; -.
CC PIR: S05008; S05008.
CC DR HSP: P00763; IDPO.
CC DR MEROPS: S01.193; -.
CC DR InterPro: IPR000152; Asx_hydroxyl.
CC DR InterPro: IPR000859; CUB_domain.
CC DR InterPro: IPR001314; Chymotrypsin.
CC DR InterPro: IPR000561; EGF-like.
CC DR InterPro: IPR001881; EGF_Ca.
CC DR InterPro: IPR001254; Ser_protease_Try.
CC DR Pfam: PF00008; EGF_1.
CC DR Pfam: PF00089; trypsin_1.
CC DR Pfam: PF00431; CUB_2.
CC DR PRINTS: PR00722; CHYMOTRYPSIN.
CC DR SMART: SM00032; CCP_2.
CC DR SMART: SM00042; CUB_2.
CC DR SMART: SM00179; EGF_CA_1.
CC DR SMART: SM00020; TRYPSIN_1.
CC DR PROSITE: PS00010; ASX_HYDROXYL_1.
CC DR PROSITE: PS01180; CUB_2.
CC DR PROSITE: PS00022; EGF_1; FALSE_NEG.
CC DR PROSITE: PS01186; EGF_2; FALSE_NEG.
CC DR PROSITE: PS01187; EGF_CA_1.
CC DR PROSITE: PS00240; TRYPSIN_DOM_1.
CC DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
CC DR PROSITE: PS00135; TRYPSIN_SER_1.
CC KM Hydrolyase; Serine proteinase; Calcium-binding; Extracellular matrix;
CC KW Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation; Glycoprotein;
CC Signal; EGF-like domain; Repeat.
CC FT SIGNAL 1 21
CC FT CHAIN 22 695
CC FT CHAIN 22 444
CC FT CHAIN 22 445
CC FT CHAIN 445 695
CC FT DOMAIN 22 124
CC FT DOMAIN 22 136
CC FT DOMAIN 181 296
CC FT DOMAIN 297 349
CC FT DOMAIN 352 415
CC FT DOMAIN 445 695
CC FT DOMAIN 137 178
CC FT MOD_RES 135 155
CC FT ACT_SITE 482 482
CC FT ACT_SITE 536 536
CC FT ACT_SITE 638 638
CC FT DISULFID 141 153
CC FT DISULFID 149 162
CC FT DISULFID 164 177
CC FT CARBOHYD 180 180
CC FT CARBOHYD 413 413

```

Query Match	33.18	Score 1244.5	DB 1	Length 695
Best Local Similarity	39.98	Pred. No. 5,7e-86		
Matches 271	Conservative 98	Mismatches 269	Indels 41	Gaps
OY	24	EP-VGRLASGFGFGEVANDDERMTLTAPGYRLRYFTFHDELSHLEVPDKLSSG	82	
Db	22	EPTMIGELSPNYQOAVPNEMEKTMWLEVPGEFVRLYFTFHDMHLESENCEYDSVQIISG	81	
OY	83	AKVATLTCGQESTTERAPAKDTYLSLQSSLDITFRSDYSNEKPFTEAPYAAEDIDEC	142	
Db	82	GVEBERLTCGQRTSKNANSPVIEEFOIPYKNLOVIFRSDFSENEEFTEGFAAYVIAIDNEC	141	
OY	143	QVANGAEATCOHCHNHNGFYCSCRAGVYLHNRKPRCSALCSQVFTQSGELSSPEP	202	
Db	142	TDFT-DV-CGHFCNNFTGGIFGFCSCPPREYFLHDMRMCQVNCSSNTTALIGELSSNP	199	
OY	203	RPYRLSSCTYSISLEEGFVILDFV-ESFDVENTHEPTLCPYDLKIQTDREHEHPCGK	261	
Db	260	GFPGPLTETHSNLTLDVIFQDTLEQKGMKRLRHGDPICPKBIT-ANSVMAPEKAYV	318	
OY	320	LKDSFISFCGEYELLOGLPLKSFYAVCKDGSMDRPMRACSIYDCGPPDLPSCGVEY	379	
Db	319	FKDYKISCVDGFEVBEVBNVSTFFYTCQSGNQMSSRLRCQYVDGITEPIQNGKVD-	377	
OY	380	ITGPGVYTKAVIYQSCETFTYTK--VNDKIVYCEADGEWTSKGEKSLVCEPVCGLS	437	
Db	378	--DENTLFGSVIHYSCEEPYVMEHAEHGEYRCANGSVNDELTELPKCPVYCGVP	435	
OY	438	AR--TTGRIYGGKAKRGDPRMVOVLIGCTTAGALLYDMWVLTAAHAYEOKHDS--	493	
Db	436	TEPRIRQIRITGEGPPAKIQSTPMQY-FEFPRAAGALIGEHWVLTAAHV-EGNSDPSMT	493	
OY	494	--ALDIRMGLKRLSPHYTQAMSEAVFIHEGY-----THDAGEINDIALIKLNKVVIN	545	
Db	494	VGSTSVREMENTANVO---KLTTTRVILLHPMKRGDDLSTPTFNDNDIALVRLKDPYKMG	549	
OY	546	SNIPRIPCRLKEAESPMTDITGASMGGLORCFELARNLMYVDIPLVDHQCTAALEKP	605	
Db	550	PVTSVTLCPGSSSEYBESGDLGLISGGRTERKRNIIYQLGAKLPLVTSLECKQVKEEN	609	
OY	606	PYPRGS--VTANMLCAGLESKGKDSGRGSGG--ALVFIDSETERFVFGVISMGSMMC	660	
Db	610	PKARADVDVFTSMNICAG--EKGVDSQGGGGAFFALPVPVRPVRKRYVAGIVSMGK---	664	
OY	661	GEAGQCYGTYTVIVYIWI 679		
Db	665	KCGYGYGTYTVKKNYKDWI 682		

RT "Humanins for complement components C1r and C1s in a close
RT tail-to-tail arrangement.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7307-7311(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88082788; PubMed=3500856;
RT McInnon C.M., Carter P.E., Smyth S.J., Dunbar B., Fothergill J.E.;
RT "Molecular cloning of cDNA for human complement component C1s. The
RT complete amino acid sequence.";
RL Eur. J. Biochem. 169:547-553(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=88163522; PubMed=2831944;
RX Tosi M., Duponchel C., Meo T., Julier C.;
RT "Complete cDNA sequence of human complement C1s and close physical
RT linkage of the homologous genes C1s and C1r.";
RL Biochemistry 26:8516-8524(1987).
RN [4]
RP SEQUENCE OF 1-329 FROM N.A.
RC TISSUE=Peripherl blood leukocytes;
RX MEDLINE=99008558; PubMed=9794427;
RA Enno Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
RA Nonaka M., Fujita T.;
RT "Two lineages of mannose-binding lectin-associated serine protease
RT (MASP) in vertebrates.";
RL J. Immunol. 161:4924-4930(1998).
RN [5]
RP SEQUENCE OF 291-688 FROM N.A.
RC MEDLINE=90040704; PubMed=2553984;
RX Tosi M., Duponchel C., Meo T., Couture-Tosi E.;
RT "Complement genes C1r and C1s feature an intronless serine protease
RT domain closely related to haploglobin.";
RL J. Mol. Biol. 208:709-714(1989).
RN [6]
RP SEQUENCE OF 16-61, 168-219, 287-334 AND 384-445.
RC MEDLINE=86164350; PubMed=3007145;
RX Spycher S.E., Nick H., Rickli E.E.;
RT "Human complement component C1s. Partial sequence determination of
RT the heavy chain and identification of the peptide bond cleaved during
RT activation.";
RL Eur. J. Biochem. 156:49-57(1986).
RN [7]
RP SEQUENCE OF 438-500, 503-534, 542-601; 617-623 AND 626-656.
RX MEDLINE=84104122; PubMed=6562661;
RA Carter P.E., Dunbar B., Fothergill J.E.;
RT "The serine proteinase chain of human complement component C1s.
RT Cyanogen bromide cleavage and N-terminal sequences of the
RT fragments.";
RL Biochem. J. 215:565-571(1983).
RN [8]
RP PARTIAL SEQUENCE.
RC TISSUE=Plasma;
RX MEDLINE=91308095; PubMed=1854725;
RA Ily C., Thieles N.M., Gagnon J., Arlaud G.J.;
RT "Effect of laccoperoxidase-catalyzed iodination on the Ca(2+)-
RT dependent interactions of human C1s. Location of the iodination
RT sites.";
RL Biochemistry 30:7135-7141(1991).
RN [9]
RP DISULFIDE BONDS.
RC MEDLINE=91157525; PubMed=2007122;
RX Hess D., Schaller J., Rickli E.E.;
RT "Identification of the disulfide bonds of human complement C1s.";
RL Biochemistry 30:2827-2833(1991).
RN [10]
RP PARTIAL SEQUENCE, AND 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.
RX MEDLINE=95288736; PubMed=7779774;
RA Rossi V., Gabotiad C., Lacroix M., Ulrich J., Fontcilla-Camps J.C.,
RA Gagnon J., Arlaud G.J.;
RT "Structure of the catalytic region of human complement protease C1s:
RT study by chemical cross-linking and three-dimensional homology
RT modeling.";

RL Biochemistry 34:7311-7321(1995).
RN [11]
RP DISEASE.
RX PubMed=11390518:
RA Dragon-Direy M.-A., Quartier P., Fremeaux-Bacchi V., Blouin J.,
de Barace C., Plier A.-M., Weiss L., Fridman W.-H.:
RT "Molecular basis of a selective C1s deficiency associated with early
onset multiple autoimmune diseases.";
RL J. Immunol. 166:7612-7616(2001).
CC -1- FUNCTION: C1S B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1Q
AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
OF THE COMPLEMENT SYSTEM. C1R ACTIVATES C1S SO THAT IT CAN, IN
TURN, ACTIVATE C2 AND C4.
CC -1- CATALYTIC ACTIVITY: CLEAVES COMPONENT C4 TO C4A AND C4B, AND
COMPONENT C2 TO C2A AND C2B.
CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
AND S IN THE MOLAR RATION OF 1:2:2. ACTIVATED C1S IS AN DISULFIDE-
LINKED HETERODIMER OF AN HEAVY CHAIN AND A LIGHT CHAIN.
CC -1- DISEASE: Defects in C1S are the cause of selective C1s deficiency,
that is associated with early onset multiple autoimmune diseases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X06596; CAA29817.1; -;
DR EMBL: J04080; AAA51852.1; -;
DR EMBL: M18767; AAA51853.1; -;
DR EMBL: AB009076; BAA86864.1; -;
DR PIR: A40496; C1HUS.
DR PIR: S00224; S00224.
DR MEROPS: S01.193; -;
DR SWISS-2DPAGE: P09871; HUMAN.
DR Genew: HGNC:1247; C1S.
DR MIM: 120580; -;
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001234; Ser_protease_Try.
DR InterPro: IPR000436; Sushl_SCR_CCP.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00089; trypsin_1.
DR Pfam: PF00431; CUB_2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP_2.
DR SMART: SM00042; CUB_2.
DR SMART: SM00179; EGF_CA_1.
DR SMART: SM00020; Tryp_SPC_1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS01180; CUB_2; 2.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
DR PROSITE: PS01187; EGF_CA_1; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Complement pathway; Plasma; Glycoprotein; Serine protease; Hydrolase;
KW Hydroxylation; Sushl; Repeat; Signal; EGF-like domain;
KW Calcium-binding.
FT SIGNAL 1 15
FT CHAIN 16 437 COMPLEMENT C1S HEAVY CHAIN.
FT CHAIN 438 688 COMPLEMENT C1S LIGHT CHAIN.

FT DOMAIN 16 130 CUB 1.
FT DOMAIN 131 172 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 175 290 CUB 2.
FT DOMAIN 293 355 SUSHI 1.
FT DOMAIN 358 422 SUSHI 2.
FT DOMAIN 438 688 SERINE PROTEASE.
FT ACT_SITE 475 475 CHARGE RELAY SYSTEM.
FT ACT_SITE 529 529 CHARGE RELAY SYSTEM.
FT ACT_SITE 632 632 CHARGE RELAY SYSTEM.
FT MOD_RES 149 149 HYDROXYLATION (PROBABLE).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .).
FT DISULFID 135 147
FT DISULFID 143 156
FT DISULFID 158 171
FT DISULFID 175 202
FT DISULFID 234 251
FT DISULFID 294 341
FT DISULFID 321 354
FT DISULFID 359 403
FT DISULFID 386 421
FT DISULFID 425 549
FT DISULFID 595 618
FT DISULFID 628 659
FT CONFLICT 294 294
FT CONFLICT 513 513
FT CONFLICT 573 573
FT CONFLICT 645 646
SQ SOURCE 688 AA; 76684 MW; 8552267A/C4C47205 CRC64;

C -> K (IN REF. 6).
G -> G (IN REF. 5).
T -> A (IN REF. 7).
TK -> GR (IN REF. 7).

INTERCHAIN.

Query Match 32.9%; Score 1238; DB 1; Length 688;
Best Local Similarity 38.2%; Pred. No. 1.7e-85;

Matches 268; Conservative 102; Mismatches 283; Indels 48; Gaps 18;

QY .1 MRLTLGLGCGSVATPLGPKWPEVGRSLASPGFPEYANDERRMTLPAPGYRLRLY 60
DB 1 MNCIVLPSLLAWVAEP-----TWYGEILSPYPAQAYSEVSEKSMIDIEVEGYHL 53
QY 61 FTHDLELSHLCEDYFVKLSGAKVLTTCQSESTDERAPGKTFYSLGSSLDITRRSD 120
DB 54 FTHLDIELSENCAYDSVQIISGDEGRGCRSSNPHSYVEFQVYNNKLVQVIFKSD 113
QY 121 YSNKPTGFEAFYAADIDECQVAPGEAPTCDHCHNHGGFYCSCRAGVLRNKRKTC 180
DB 114 FSNERTGFAYIVADINCTDFV-DVP-CSHFNNFTGGFSCPPREYFLHDDMKNC 171
QY 181 SALCSGVFTQSGELSPPEYPRPKLSCTYSISLEEGFVILDF-VESFDVETHPET 239
DB 172 GVCNSGVFTALIGELIASPNVPRKPYSPNSRCEYIRLEKGFQVAVTLRREDFVEADSA 231
QY 240 LCPYDFLKIQTDRREHRCFPGCKTLPH--RIETKSNVTYITTFVWDESGDHTGKIHVSTA 297
DB 232 GNCCLDSLVFAVAGDRQFGPGCHGFRPLNIETKSNALDIITQDILTLQCKGKMKLRVGGDP 291
QY 298 QPCPYMAPRNGHVSPOAKYILKDSFICETGYELLOGHLPLKSPFAVCOKGSMDRP 357
DB 292 MPCP-KEDTPNSVPRKAKIVPRDVOITCLDGFVEYEVGRVGTSTYSCQSGKKSNS 350
QY 358 MPACSIIVDCGPPDDLPSPGRVEYITGPGVTTYKAVIYQSCSETFYTMK-VNDGKYVCBAD 416
DB 351 KLCQAPVDCGIPESIEENGKVE---DPRESTLEGVIRVTCPEPPYVMENGGGGEVHCAGNG 407
QY 417 FWTSSKEKSLPVCPEPYCGLSAR--TTGGRIYGGQAKKPDGPPQVUILLGTTAAGLLY 474
DB 408 SWNEVLGPRLPKCVPGVPRPEFEKORLIGSDSDIKNFPMQV-FPDNPMAGGALIN 466
QY 475 DMNVLTAAHAYVEQK-----DASALDIRNGTLKRLSPHYTQAMSPAVFTHEGY----- 523
DB 467 EYVLTAAHAYVEGNREPTMYVGSVTSRLAKSKMLTP-----EAVFIHQGKMLEV 519
QY 524 -THDAGFDNDIALIKLNKKVINSNIPICLPKKEASFMTDDIGTASGWLTORGFALA 582
DB 524 -THDAGFDNDIALIKLNKKVINSNIPICLPKKEASFMTDDIGTASGWLTORGFALA 582

Db 520 PEGRTNEDNIALVRLKDPVKGFTVSPICLPGTSSDYNLMDGLISGWRTEKRDRA 579

QY 583 RNLMVVDIPIDHOKC-TAAEKPEPYRGS--VTANMTCAGLESQKSGDGGALVF 639

Db 580 VRLKARLPVAPLRCKCKVKEKFPADAEAYFTPMTCAGGCK-GMDSCKDGGAGAV 638

QY 640 LD-SETERMFVGIVSGSNMGAGGYVYTKYINYPWI 679

Db 639 QDPNDKTKFYAAGLVSMGP-----QCGTYGLTFRVKNYDWI 675

RESULT 7

ST14_HUMAN STANDARD; PRT: 855 AA.

ID ST14_HUMAN

AC 09Y5Y6: 09H350: 09HCA3: 09BS01: 09HB36;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DR Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrilysin) (Membrane-type serine protease 1) (MT-SPI) (Prostamin) (Serine protease TADG-15) (Tumor associated differentially-expressed gene-15 protein).

ST14 OR PRSS14 OR SMC19 OR TADG15.

CC Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-99303581; PubMed-10373424;

RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;

RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine protease with trypsin-like activity.";

RL J. Biol. Chem. 274:18231-18236(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-99432178; PubMed-10500122;

RA Takeuchi T., Shuman M.A., Craik C.S.;

RT "Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).

RN [3]

RP SEQUENCE FROM N.A.

RX TISSUE-Prostate;

RA Yamaguchi N., Mitsui S.;

RT "Molecular cloning of a novel transmembrane serine protease expressed in human prostate.";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RX Yamamoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H., O'Brien T.J.;

RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE OF 327-855 FROM N.A.

RX TISSUE-Muscle;

RA Strausberg R.;

RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE OF 340-664 FROM N.A.

RX Cao J., Fan W., Zheng S.;

RT "Genomic analysis of a novel human serine protease SMC19.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RN [7]

RP CHARACTERIZATION.

RX TISSUE-Milk;

RX MEDLINE-99303582; PubMed-10373425;

RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;

RT "Purification and characterization of a complex containing matrilysin and a Kunitz-type serine protease inhibitor from human milk.";

RL J. Biol. Chem. 274:18237-18242(1999).

CC -1- FUNCTION: DEGRADATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE

CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG OR LYS AS THE P1 SITE.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF118324; AAD42765.2; -

DR EMBL: AF133086; AAF00109.1; -

DR EMBL: AB030036; BAB20376.1; -

DR EMBL: AF057145; AAG15395.1; -

DR EMBL: BC005826; AAH05826.1; -

DR EMBL: AF283256; AAG13949.1; -

DR HSSP: P00763; 1DPO. 34.

DR Genew: HGNC:11344; ST14.

DR MIM: 606797; -

DR MEROPS: S01.302; -

DR InterPro: IPR000859; CUB_domain.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR002172; LDL_receptl_A.

DR InterPro: IPR001254; Ser_protease_Try.

DR Pfam: PF00057; Idl_receptl_a; 4.

DR Pfam: PF00089; trypsin; 1.

DR Pfam: PF00431; CUB; 2.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS: PR00261; LDLRECEPTOR.

DR SMART: SM00042; CUB; 2.

DR SMART: SM00192; LDLA; 3.

DR SMART: SM00020; TRYP_SPC; 1.

DR PROSITE: PS01180; CUB; 2.

DR PROSITE: PS01209; LDLRA_1; 2.

DR PROSITE: PS50068; LDLRA_2; 4.

DR PROSITE: PS50240; TRYPsin_DOM; 1.

DR PROSITE: PS00134; TRYPsin_HIS; 1.

DR PROSITE: PS00135; TRYPsin_SER; 1.

KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;

KW Transmembrane; Repeat.

FT DOMAIN 1 55

FT TRANSMEM 56 76

FT 77 855

FT DOMAIN 214 334

FT DOMAIN 340 447

FT DOMAIN 452 487

FT DOMAIN 487 524

FT DOMAIN 524 560

FT DOMAIN 566 603

FT DOMAIN 615 854

FT ACT_SITE 656 656

FT ACT_SITE 711 711

FT ACT_SITE 805 805

FT CARBOHYD 109 109

FT CARBOHYD 302 302

FT CARBOHYD 485 485

FT CARBOHYD 772 772

FT CONFLICT 327 329

FT CONFLICT 381 381

FT CONFLICT 674 674

FT SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 13.0%; Score 488; DB 1; Length 855;

Best Local Similarity 25.5%; Pred. No. 5.7e-29;

Matches 186; Conservative 90; Mismatches 270; Indels 184; Gaps 31;

QY 29 RLASGFP-GEYANDQERRWTLTAPGYSRLRYFTHEDELSHLCEYDFVKSAGAK--- 84

```

Db 227 RFTTFGPDSEYPAHARQOMALRQDADSVLSTFSPFLASCDESGSLVTVYNTLSPME 286
QY 85 --VLATLQGEOSTDERAPAGKDTFYSGLSSDITFRSDYSNEKPREPTGEA-FYAADIDE 141
Db 287 PHALVOLCG-----TTPSYMLTFRSSONVLLITLITERRRH--GEAATFOLPRMS 339
QY 142 QOVAPGEAPTCDHCHNHLLGFGYCSGRAGVYLHRNKRTCSALCSGVFTORSSELSPREY 201
Db 340 C-----GRLRKAGQGTFFNSPY 356
QY 202 PRPYKLSCTYSISLSEGEVILDFVSEDEVTH-PETLCPYDPLKIQTDREHGPRCG 260
Db 357 PGHYPNPIDCTWNIIEVPNNQHVKKRFFYLEPEVPAGTGPCKDVEINGEK-----YCG 411
QY 261 KTLPHRIETGNTFYITFVTDESQDTHKTHYTS--TAQCCPYMAPNPGVSPVQAKY 318
Db 412 ERSGEVVYSNKNKTIKTRHSDQSTIDTGFLLAYLSYDSSDFCP-----GQFTCTRCR 464
QY 319 ILK-----DSFISFETGYELLQGLPLKSFYAVCOK-----DGSMDRPM 359
Db 465 IRKELRCQDMADCTDHSDELNCSDAGHQFTCKNFKPLFMVCOVSNDGDNDEQ--- 521
QY 360 ACSIV-----DCGP-PDDLPGSRVEYITGPGVTTYKAVIQYSC 396
Db 522 GCSGPAQTFRCNSGKCLSKSOQCNKDKDGDGSDSEASCPKVVVYCTKHT-----YRC 574
QY 397 EETFTYTMKN---DQKYVCEADGFWTSKSGKSLPVCPEVGLSARTGSGRIYGGQKAKP 453
Db 575 LNLGLCLSGNNECDKEK-SDG-----SDEKD--CD--GLKNSFTQARVAGGTDADE 623
QY 454 GDFPMOVLII--LG-GTTAAGALLVDMVLTAAHAYEOK-----HD--- 491
Db 624 GEMPMOVSALHAGGHCIGASLISPNMLVSAACHYIDDRGRYSPTQMTAFGLHDSQ 683
QY 492 ASALDIRGTIKRLSPHTQAMSENVFHEGYTHDAGDNDALIKLNKKVYINSNTPI 551
Db 684 RSAPQGVQERRKRIISH-----PFENDFPTDYDIALLEKPKPAEYSMVAPI 730
QY 552 CLPKRAESFMRTDIGTASGMGLTORGFL-ARNLMYVIDIPDHQKTAAYEKPPYRG 610
Db 731 CLP---DASHVPRPAGKAIWYTMGHQYGTGALLQKEIRVINYNTTENLPPQ----- 782
QY 611 SVTANMLCAGLESGKSDSGALVLDSETRMFVGVSGVSNMCGEAGQGYVT 670
Db 783 QITPRMCMVGLSGVDSCQDSDGSLSSVEAD-GRIFQAGVVSNGD-GCAQRMKPGYIT 840
671 KVINIPIE 680
841 RLPLERDMIK 850

```

```

CC -1- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
CC SENSITIVE HEMOLYMPH COAGULATION SYSTEM WHICH MAY PLAY IMPORTANT
CC ROLES IN BOTH HEMOSTASIS AND HOST DEFENSE MECHANISMS. ITS ACTIVE
CC FORM CATALYZES THE ACTIVATION OF FACTOR B.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 103-Arg-1-Ser-104 and
CC 124-Ile-1-Ile-125 bonds in Limulus clotting factor B to form
CC activated factor B. Cleavage of pro-Arg-1-Xaa bonds in synthetic
CC substrates.
CC -1- ENZYME REGULATION: ACTIVATED BY GRAM-NEGATIVE BACTERIAL
CC LIPOLYSACCHARIDES AND CHYMOTRYPSIN.
CC -1- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90271; BAAL4315.1; -.
CC EMBL: D90272; BAAL4316.1; -.
CC PIR: A38738; A38738.
CC PIR: B38738; B38738.
CC HSSP: P00763; IDPO.
CC MEROPS: S01.219; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR004043; LCCL_dom.
CC InterPro: IPR001304; Lectin_C.
CC InterPro: IPR001254; Ser_protease_Try.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: PF00059; Lectin_C.1.
CC Pfam: PF00084; Sushi.5.
CC Pfam: PF00089; Trypsin.1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00032; CCP.5.
CC SMART: SM00034; CLECT.1.
CC SMART: SM00181; EGF.1.
CC SMART: SM00020; Tryp_Spc.1.
CC PROSITE: PS00615; C-TYPE_LECTIN_1; FALSE_NEG.
CC PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; FALSE_NEG.
CC PROSITE: PS50820; LCCL.1.
CC PROSITE: PS50240; TRYPSIN_DOM.1.
CC PROSITE: PS00134; TRYPSIN_HIS.1.
CC PROSITE: PS00135; TRYPSIN_SER.1.
CC Hydrolyase: Serine protease; Signal: Alternative splicing; Lectin;
CC Hemolymph clotting; Glycoprotein; Cell adhesion; EGF-like domain;
CC Sushi; Repeat.1
CC SIGNAL 1 25
CC CHAIN 26 1019
CC CHAIN 26 690
CC CHAIN 691 1019
CC CHAIN 691 762
CC CHAIN 763 1019
CC DOMAIN 102 137
CC DOMAIN 142 195
CC DOMAIN 200 254
CC DOMAIN 321 421
CC DOMAIN 436 568
CC DOMAIN 576 634
CC DOMAIN 685 748
CC DOMAIN 763 1019
CC ACT_SITE 809

```

```

LIMULUS CLOTTING FACTOR C.
LIMULUS CLOTTING FACTOR C, HEAVY CHAIN.
LIMULUS CLOTTING FACTOR C, LIGHT CHAIN.
LIMULUS CLOTTING FACTOR C, A CHAIN.
LIMULUS CLOTTING FACTOR C, B CHAIN.
EGF-LIKE.
SUSHI 1.
SUSHI 2.
SUSHI 3.
C-TYPE LECTIN.
SUSHI 4.
SUSHI 5.
SERINE PROTEASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT ACT_SITE 865 865 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 966 966 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT BINDING 960 960 SUBSTRATE (BY SIMILARITY).
FT DOMAIN 643 689 PRO-RICH.
FT DISULFID 106 118 BY SIMILARITY.
FT DISULFID 112 125 BY SIMILARITY.
FT DISULFID 127 136 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 464 564 BY SIMILARITY.
FT DISULFID 538 556 BY SIMILARITY.
FT DISULFID 794 810 BY SIMILARITY.
FT DISULFID 932 951 BY SIMILARITY.
FT DISULFID 962 996 BY SIMILARITY.
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 740 740 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 767 767 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 912 912 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 492 498 LTTWIG -> TDWYAT (IN SHORT ISOFORM).
FT VARSPLIC 499 1019 MISSING (IN SHORT ISOFORM).
FT SEQUENCE 1019 AA; 112346 MW; 5BC2864C6715289B CRC64;

Query Match 12.5%; Score 471.5; DB 1; Length 1019;
Best Local Similarity 28.4%; Pred. No. 1.2e-27;
Matches 152; Conservative 61; Mismatches 190; Indels 133; Gaps 27;

QY 265 HRIEF-----KSMTV---TTF-VTDESGDHTG-----WKIHYSTA 297
DB 500 HRLDEKPEFWELMDRNSVNLNLTFWASGEPNETCYLIDRIDLOPWKT--KSCF 557
QY 298 QP-----CPYMAPPNNGHVSVQAKYILKDSIFCEYELLGHLPL 341
DB 558 QPSSPACAMDLSDRKAKACDDEPLENGHAT-LHGQSI--DGFYAGSSIRSCVELHLIS 614
QY 342 KSFYAVCKQKDSWDRPMAC-SIYDC-GPPDDLPS-GRVEYITGCVTT----- 387
DB 615 GTEVTCTTNGTWSAPKPRCIVITCOMPF--VPSYGSVE--IKPPSRNISIRVSGPELR 671
QY 388 -----YKAVIO--YSCEEFFYMKYND 407
DB 672 LPRLEPLPRAAKAPPRKRRSSOPSTVDLASKVKLPBGHRRVGSRAIVYCESRYVELLSQ 731
QY 408 GKYYCEADGEMTTSKGEKSLPVCPEVCGLSARTTGRTYGGOKAKGDFPMQVLI----- 462
DB 722 GRR-CDNSGNMNSGRPAS-----CIYVCGRSDSPRSPIFMNGSTPIGQPMQAGISRWLA 785
QY 463 ---LGGTTAAGALYDNNVLTAAHAY---YEQKHDSALDIRMGTLKR---LSPHYTQA 512
DB 786 DHNMWFLCCGGSLANEKWIYTAHCVTSATVAEIIDPSQFIYILCKYVRDSDRDYQV 845
QY 513 MSEAVFIHEGYTHDAGFON-DIALIKLNNKVINSNIPICLPKR-EAESFRTRDICTA 570
DB 846 -REALIEIHNPNYDPGNINFDALILQLTQTPVLTTRVPICLPDITTRREHKLETLAVV 904
QY 571 SGWGLTQGFGLARNLMLYVDIPVHOKCTAAYEKPPRGSVTANMLCAGLESCKDSCR 630
DB 905 TGMVLENNTYSEMIQQAIVLPVVAASSTCEBQKEADLPL-ITVEMFPAGAKKKGKGYDCS 963
QY 631 GDSGALVFL-DSETE-RWFLVGGIYSWGS-MNCGEAGYGYTKYINIPWIENTII 683
DB 964 GDSGPLVFLFADSDRTERRMVLEGIVSMGSCCKRANQYGGFTTKVNVFLSWIROFI 1019

RESULT 9
LFC_CARRO STANDARD; PRT; 1019 AA.
AC Q26422;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Limulus clotting factor C precursor (EC 3.4.21.84) (FC).
OS Carinoscorpius rotundicauda (Southeast Asian horseshoe crab).

```

```

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Carinoscorpius.
OC NCBI_TaxID=6848;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=95268506; PubMed=7538401;
RA Ding J.L., Navas M.A. III, Ho B.;
RT "Molecular cloning and sequence analysis of factor C cDNA from the
RT Singapore horseshoe crab, Carinoscorpius rotundicauda.";
RL Mol. Mar. Biol. Biotechnol. 4:90-103(1995).
CC -1- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
CC SENSITIVE HEMOLYMPH COAGULATION SYSTEM WHICH MAY PLAY IMPORTANT
CC ROLES IN BOTH HEMOSTASIS AND HOST DEFENSE MECHANISMS. ITS ACTIVE
CC FORM CATALYZES THE ACTIVATION OF FACTOR B.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 103-Arg-1-Ser-104 and
CC 124-Ile-1-Ile-125 bonds in Limulus clotting factor B to form
CC activated factor B. Cleavage of Pro-Arg-1-Xaa bonds in synthetic
CC substrates.
CC -1- ENZYME REGULATION: ACTIVATED BY GRAM-NEGATIVE BACTERIAL
CC LIPOLYSACCHARIDES AND CHYMOTRYPSIN (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S77063; AAB34361.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.219; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR004043; LCCL-dom.
DR InterPro; IPR001304; LECTIN_C.
DR InterPro; IPR001254; Ser protease_Try.
DR Pfam; PF00059; LECTIN_C; 1.
DR Pfam; PF00084; Sush1; 5.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00032; CCP; 5.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; FALSE_NEG.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS50820; LCCL; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Signal; Lectin; Hemolymph clotting;
KW Glycoprotein; Cell adhesion; EGF-like domain; Sush1; Repeat.
FT SIGNAL 1
FT CHAIN 1
FT CHAIN 26 1019 LIMULUS CLOTTING FACTOR C.
FT CHAIN 26 690 LIMULUS CLOTTING FACTOR C.
FT CHAIN 691 1019 LIMULUS CLOTTING FACTOR C.
FT CHAIN 691 762 LIMULUS CLOTTING FACTOR C.
FT CHAIN 763 1019 LIMULUS CLOTTING FACTOR C.
FT DOMAIN 102 137 EGF-LIKE.
FT DOMAIN 142 195 SUSHI 1.
FT DOMAIN 200 254 SUSHI 2.

```

```

FT DOMAIN 260 321 SUSHI 3.
FT DOMAIN 325 421 LCCL.
FT DOMAIN 436 568 C-TYPE LECTIN.
FT DOMAIN 576 634 SUSHI 4.
FT DOMAIN 685 748 SUSHI 5.
FT DOMAIN 763 1019 SERINE PROTEASE.
FT ACT_SITE 809 809 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 865 865 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 966 966 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT BINDING 960 960 SUBSTRATE (BY SIMILARITY).
FT DOMAIN 643 689 PRO-RICH.
FT DISULFID 106 118 BY SIMILARITY.
FT DISULFID 112 125 BY SIMILARITY.
FT DISULFID 127 136 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 464 564 BY SIMILARITY.
FT DISULFID 538 556 BY SIMILARITY.
FT DISULFID 794 810 BY SIMILARITY.
FT DISULFID 932 951 BY SIMILARITY.
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 740 740 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 767 767 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 912 912 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1019 AA: 112429 MW: 918A1ED8B817B6C3 CRC64:

```

```

Query Match 12.5% Score 470.5; DB 1; Length 1019;
Best Local Similarity 28.1% Pred. No. 1.5e-27;
Matches 151; Conservative 63; Mismatches 187; Indels 137; Gaps 26;

```

```

QY 265 HRIET-----KSNV---TTF-VTDESQHTG-----WKIHYSTA 297
DB 500 HRLDAEKFTELMDRBNVNLDTFMASGEPEENETNCVMDIQDQOSWTK--KSCF 557
QY 298 QP-----CPYPAHPNGHVSPOAKYILNDSIESIFETGYELQHLPL 341
DB 558 QPSSFAECMDLSDRNKAKCDPGLENGHAT-LHGOSI--DGFAGSSIRSCVLYHLS 614
QY 342 KSFTAVCQKDGSMRPPAC-SIVDC-GPPDDLP-S-GRVETITPGVYT-----387
DB 615 GTEIVTCTTGTSAPKPRCIVKTYTCNPP--VPSIGSVE-IRPPSRNISRVSPPPLR 671
QY 388 -----YKAVID--YSCETFTMKVND 407
DB 672 LPLRLPLARAKPPKPRSSQSTVDLASKVKLPEGHYRGSRAIYTCESRYELLGSQ 731
DB 408 GKYYCEADGFWTSKSGKSLPVCPEPVGLSARTTGGRIYGGQAKPDGFWQVLI-----462
DB 732 GRR-CDSENGMNSGRPAS-----CIPVCGRSDSPSPFIMNNGNSTEIGQMPQACISRWLA 785
QY 463 ---LGTTAAGALLYDMVVLTAHAHVEOKHDAALDIRMGTILKRLSPHYQAMS-----514
DB 786 DHNMWFLQCGGSLNENKMYIAHCV---TYSATAEIIDPNQFMWYLGKYYRDSRDDY 842
QY 515 ---EAVFIHEGYTHDAFDN-DIALIKLNKKVYINSNITPICLPK-FAESFRTDDIG 568
DB 843 VQVEALEIHVNPYDGNLNFIDALQLKTPVLTTRVQICLPDITTRHLEKEGTIA 902
QY 569 TASGWTGQKGLARNLMTYVDIPVDHOKTAAYEKPPYPRGSVYANMLCAGLESKDS 628
DB 903 VVVGWGLNENNTYSETIQOAVLPVVAASCEEGYKEADLPL-TVTENNFCAGYKKGRYDA 961
QY 629 CRQSGGALVFL-DSETE-RMFVGGIVSMGS-MNCGEAGQGVYTKYINYPMENTII 683
DB 962 CSGDSGGLPVFADDSRTERRVLEGIIVSMGSPSCGKANOYGGTFKAVNVLFWITROPI 1019

```

```

RESULT 10
ST14 MOUSE
ID ST14 MOUSE STANDARD: PRT: 855 AA.
AC P56677;

```

```

DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithlin).
GN ST14 OR PRSS14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RX MEDLINE=99216440; PubMed=10199918;
RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA Schwartz R.H.;
RT "Cloning and chromosomal mapping of a gene isolated from thymic
RT stromal cells encoding a new mouse type II membrane serine protease,
RT epithlin, containing four LDL receptor modules and two CUB domains.";
RL Immunogenetics 49:420-428(1999).
RN [2]
RP REVISIONS TO 23: 321; 325; 343; 409-410 AND C-TERMINUS.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (probable).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART.
CC TESTIS, AND BRAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: AF042822; AAD02230.3; -
DB EMBL: BC005496; AA05496.1; -
DB HSP: P20231; IAAO.
DB MEROPS: S01.302; -.
DB MGD: MGI:1338881; St14.
DB InterPro: IPR000859; CUB domain.
DB InterPro: IPR002172; LDL_receptl_A.
DB InterPro: IPR001254; Ser_protease_Try.
DB Pfam: PF00057; Idl_receptl_a; 4.
DB Pfam: PF00089; trypsin; 1.
DB SMART: SM00042; CUB; 2.
DB SMART: SM00192; LDLa; 3.
DB PROSITE: PS01180; CUB; 2.
DB PROSITE: PS01209; LDLRa_1; 2.
DB PROSITE: PS50068; LDLRa_2; 4.
DB PROSITE: PS50240; TRYPsin_DOM; 1.
DB PROSITE: PS00134; TRYPsin_HIS; 1.
DB PROSITE: PS00135; TRYPsin_SER; 1.
KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KW Transmembrane; Repeat.
FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 214 331 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 340 444 CUB 1.
FT DOMAIN 451 488 CUB 2.
FT LDL-RECEPTOR CLASS A 1.

```



```

FT DOMAIN 125 168 EGF-LIKE 2.
FT DOMAIN 241 475 SERINE PROTEASE.
FT MOD_RES 46 46 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 47 47 (BY SIMILARITY)
FT MOD_RES 47 47 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 54 54 (BY SIMILARITY)
FT MOD_RES 54 54 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 56 56 (BY SIMILARITY)
FT MOD_RES 56 56 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 59 59 (BY SIMILARITY)
FT MOD_RES 59 59 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 60 60 (BY SIMILARITY)
FT MOD_RES 60 60 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 65 65 (BY SIMILARITY)
FT MOD_RES 65 65 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 66 66 (BY SIMILARITY)
FT MOD_RES 66 66 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 69 69 (BY SIMILARITY)
FT MOD_RES 69 69 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 72 72 (BY SIMILARITY)
FT MOD_RES 72 72 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 79 79 (BY SIMILARITY)
FT MOD_RES 79 79 GAMMA-CARBOXYGLUTAMIC ACID
FT MOD_RES 103 103 (BY SIMILARITY)
FT MOD_RES 103 103 HYDROXYLATION (BY SIMILARITY).
FT ACT_SITE 282 282 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 328 328 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 425 425 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 90 101 (BY SIMILARITY)
FT DISULFID 95 110 (BY SIMILARITY)
FT DISULFID 112 121 (BY SIMILARITY)
FT DISULFID 129 140 (BY SIMILARITY)
FT DISULFID 136 152 (BY SIMILARITY)
FT DISULFID 154 167 (BY SIMILARITY)
FT DISULFID 175 348 INTERCHAIN (BY SIMILARITY).
FT DISULFID 247 252 (BY SIMILARITY)
FT DISULFID 267 283 (BY SIMILARITY)
FT DISULFID 396 410 (BY SIMILARITY)
FT DISULFID 421 449 (BY SIMILARITY)
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 475 AA; 53142 MW; 570BF84956C574D CRC64;

```

Query Match 11.28; Score 422; DB 1; Length 475;
 Best Local Similarity 23.18; Pred. No. 2.5e-24;
 Matches 143; Conservative 78; Mismatches 169; Indels 230; Gaps 26;

```

90 CGQESTTERAPGKDTFYSIGSSLDITFRSDYSNEKPFTEAFYAEDIDECVARGEA 149
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 57 CNEERCSEKEA-----REAFEDNEKTEEFWNITV--DQGQCSNP--- 94

150 PTCDH--HCHNHGLGFYSCRAG-----VLHR---NKRITCALCSGOVTFQRSGEL 196
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 95 --CHYGQCKDGLGSLGYSCLDGYQKNCFEVIRPKYCKINNGDCEQCFKSIKSKD-- 149

197 SSPEYPRPYKLSCTYSISLEBFSVILDFVESFDVETHEITCLPDLFKIQIDREHG 256
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 150 -----VVCST-----SGVELAEQCKQCVSKVKP- 174

257 PFCGKTLPHRIE-----TKSNVTITFVTDSDGHTGMKIHSTAPCPYPMAPNGH 310
  | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 175 --CGKVLAKRIKRSVILPTSNM---NATSDQVPT-----NGS 209

311 VSPQAVYIIKDSFSICETGEYELLOGLPLKSTAVCQKDGWDRMPACSIIVDCGPPD 370
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 210 -----ILEEVFTTTES-----PTP-----PPR 227

371 DLPSGRVEYITGPGVTYTKAVIOYSCSEETPYTMKVDGKYCEADGFWTSGKESLPVC 430
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 228 NGSS-----ITDPNVDT----- 239

```

```

Qy 431 EPVCGLSARTTGGRIYGGOKAKPGDPFPOVLIL---GGTTAGALLDNWVLTAAHAYE 487
  | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 240 -----RIVGDCRCPGECPPQAVLINEKEEFCGGTILNEDFLTAHACINQ 266

Qy 488 OKHDASLDIRMGTLKRLSPHYTOA--WSEAVPIHEGYTHDAGEDNDIALIKLNKVIN 545
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 287 SKE---IKVVGVEVDREKEHSETHETAEIKFVHSKYIAET--YDNDIALIKLEPIQFS 341

Qy 546 SNIPRLPKKE--AESPMRTDDIGTASGMLT--ORGFLARLNVVDIPIVDHOKCTAAE 603
  | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 342 EYVPACLPQADPANEVLMNOKSGVSGFGRFEPAGRLSKLVLPYVDRSTCKOS-- 399

Qy 604 KPPYRGSVTANMLCAGLESCKDSCRGSGALVFLDSETERMFVSGVSGMNGEA 663
  | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 400 ----TNAITENMFCAGYETDQKACGDSGCPHYTRYKDT--YFVGIYSWGF--GCARK 452

Qy 664 GQGVYTKVINYIPWIENTII 683
  | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 453 GKYGVTYKLSRFLRWRTVM 472

```

```

RESULT 13
HPT_RAT
ID HPT_RAT STANDARD; PRT; 347 AA.
AC P06866;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Haptoglobin precursor.
GN HP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
ON [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205837; PubMed=2320005;
RA Marinkovic S., Baumann H.;
RT "Structure, hormonal regulation, and identification of the
RT interleukin-6- and dexamethasone-responsive element of the rat
RT haptoglobin gene."
RL Mol. Cell. Biol. 10:1573-1583(1990).
RN [2]
RP SEQUENCE OF 76-347 FROM N.A.
RX MEDLINE=84264555; PubMed=6204979;
RA Goldstein L.A., Heath E.C.;
RT "Nucleotide sequence of rat haptoglobin cDNA. Characterization of the
RT alpha beta-subunit junction region of prohaptoglobin."
RL J. Biol. Chem. 259:9212-9217(1984).
RN [3]
RP SEQUENCE OF 19-43 AND 103-127.
RX MEDLINE=83238370; PubMed=6863267;
RA Hanley J.M., Haugen T.H., Heath E.C.;
RT "Biosynthesis and processing of rat haptoglobin."
RL J. Biol. Chem. 258:7858-7869(1983).
RN [4]
RP SEQUENCE OF 103-142.
RX MEDLINE=77025019; PubMed=975782;
RA Kurosky A., Kim H.H., Touchstone B.;
RT "Comparative sequence analysis of the N-terminal region of rat,
RT rabbit, and dog haptoglobin beta-chains."
RL Comp. Biochem. Physiol. 55B:453-459(1976).
CC -1- FUNCTION. HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN,
CC PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
CC KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
CC ACCESSIBLE TO DEGRADATIVE ENZYMES.
CC -1- SUBUNIT. TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER, THEN RELEASED
CC IN PLASMA.
CC -1- SIMILARITY: ALTHOUGH THE BETA CHAIN IS CLEARLY RELATED TO SERINE
CC PROTEASES, HAPTOGLOBIN HAS NO ENZYMATIC ACTIVITY. POSITIONS
CC CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES OF THE

```

PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPOGLOBIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M34232: AAA41348.1; AUT_SEQ.
 DR EMBL: M34230: AAA41348.1; JOINED.
 DR EMBL: M34231: AAA41348.1; JOINED.
 DR EMBL: K01933: AAA41349.1; -
 DR PIR: A20881: HPRT.
 DR PIR: A34784: A34784.
 DR HSSP: P00734: 1BA8.
 DR MEROPS: S01.972: -
 DR InterPro: IPR001314: Chymotrypsin.
 DR InterPro: IPR001254: Ser_protease_Try.
 DR InterPro: IPR000436: Sush1_SCR_CCP.
 DR Pfam: PF00084: sush1.1.
 DR Pfam: PF00089: trypsin.1.
 DR SMART: PR00722: CHYMOTRYPSIN.
 DR SMART: SM00032: CCP.1.
 DR SMART: SM00020: TRYP_SPC.1.
 DR PROSITE: PS50240: TRYPSIN_DOM.1.
 DR KMW Acute phase: Laver: Signal.

FT SIGNAL 1 18
 FT CHAIN 19 101 HAPOGLOBIN ALPHA CHAIN.
 FT CHAIN 103 347 HAPOGLOBIN BETA CHAIN.
 FT DOMAIN 103 347 SERINE PROTEASE.
 FT DISULFID 52 86 INTERCHAIN (WITH ALPHA) (BY SIMILARITY).
 FT DISULFID 90 207 INTERCHAIN (ALPHA-BETA) (BY SIMILARITY).
 FT DISULFID 250 281 BY SIMILARITY.
 FT DISULFID 292 322 BY SIMILARITY.
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 25 25 A -> D (IN REF. 3).
 FT CONFLICT 332 332 K -> R (IN REF. 2).
 SQ SEQUENCE 347 AA; 38549 MW; BF0073DCAAB705C9 CRC64;

Query Match 11.1%; Score 419; DB 1; Length 347;
 Best Local Similarity 30.1%; Pred. No. 2.8e-24;
 Matches 101; Conservative 52; Mismatches 143; Indels 40; Gaps 11;
 366 CGPPDDLPSCRVETTPGVYTYKAVIQYSCETFTYMKVNDGKYVCEADGFWTSKGEK 425
 33 CPKPEPIANGVEHL-----VRYRCRQFYKLTQEDGITYLTINSEKQVNPAAGD 81
 426 SLRPEVPCGASARTTG--GRIVGOKAKPDPEVOVLIG--CTTAAGALLYNWVLT 481
 82 KLPKCEAVCGPKRPVQVOVITIGSGMDAKSPFWQAKMISRHLTGATLTISQWLLTT 141
 482 AAHAYEOKHDAALIDIMGLTKRLSPHYT--QANSEAVFIHEGYTHADGENDIALIKLN 539
 142 AONLF-LNHSENA-----TAKDIAPLTLTVYGNQVLVEIKVVLHRSRVDIGLILKL 194
 540 NKVAVINSNTPICLPKREAESFMRTDITGASGGLTORGFLARNLWVDIPVDHOKCT 599
 195 OKVLTETKVMPICLPSKPD---YVAPGRMGYSGWRNVNFRFTRLKYVMLPVADQKEE 251
 600 AAVEKPPYP--RGSVT-----AMMLCAGLESSGKDRSGSGGALVFLDSETERWF 648
 252 LHTEKSTVPEKGAIVTVGVQPIINKHTEFCAGLTKYEEDTCYGDAGSAFAVHDEDTWY 311
 649 VGGIVSGMNCGEAGQYGVYTKYINIPWENTIS 684
 312 AAGILSF-DKSCAVA--EYGVYVKAITDLKDWQETMA 345

RESULT 14

ENTR_PIG STANDARD: PRT: 1034 AA.
 AC P98074:
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
 GN PRS57 OR ENTR.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Ductal mucosa;
 RX MEDLINE=94327548; PubMed=8051081;
 RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S.,
 RA Miki K., Kurokawa K., Tashiro K., Shiohara K., Shimomiya K.,
 RA Uneyama H., Inoue H., Takahashi T., Takahashi K.;
 RT "Structural characterization of porcine enteropeptidase.";
 RT J Biol Chem. 269:19976-19982(1994).
 CC FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-1-Ile-7 bond in
 CC trypsinogen.
 CC SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
 CC (HEAVY) CHAIN, AND A MINI CHAIN.
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC PTM: THE CHAIN IS DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
 CC SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC SIMILARITY: CONTAINS 1 SCR DOMAIN.
 CC SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D30799: BAA06459.1; -
 DR HSSP: P00763: LDPO.
 DR MEROPS: S01.156: -
 DR InterPro: IPR000859: CUB_domain.
 DR InterPro: IPR001314: Chymotrypsin.
 DR InterPro: IPR002172: LDL_recept_A.
 DR InterPro: IPR000998: MAM_domain.
 DR InterPro: IPR000082: SEA_domain.
 DR InterPro: IPR001254: Ser_protease_Try.
 DR InterPro: IPR001190: Srcf_receptor.
 DR Pfam: PF00057: Idl_recept_a.2.
 DR Pfam: PF00089: trypsin.1.
 DR Pfam: PF00431: CUB.2.
 DR Pfam: PF00530: SRCR.1.
 DR Pfam: PF00629: MAM.1.
 DR Pfam: PF01390: SEA.1.
 DR PRINTS: PR00722: CHYMOTRYPSIN.
 DR SMART: SM00042: CUB.2.
 DR SMART: SM00192: LDLa.2.
 DR SMART: SM00137: MAM.1.
 DR SMART: SM00200: SEA.1.
 DR SMART: SM00202: SR.1.
 DR SMART: SM00020: TRYP_SPC.1.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=94229342; PubMed=8174772;
 RA Fukagawa M., Noboru S., Hogan B.L.M., Jones C.M.;
 RT "Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),
 RT which is related to the Drosophila dorsoventral gene tolloid and
 RT encodes a putative astacin metalloendopeptidase";
 RL Dev. Biol. 163:175-183(1994).
 CC -1- FUNCTION: CLEAVES THE C-TERMINAL PROPEPTIDES OF PROCOLLAGEN I, II
 CC AND II. INDICES CARTILAGE AND BONE FORMATION.
 CC -1- CATALYTIC ACTIVITY: cleavage of the C-terminal propeptide at
 CC Ala-1-asp in type I and II procollagens and at Arg-1-asp in type
 CC III.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- ENZYME REGULATION: ACTIVITY IS INCREASED BY THE PROCOLLAGEN C-
 CC ENDOPEPTIDASE ENHANCER PROTEIN.
 CC -1- TISSUE SPECIFICITY: AT HIGH LEVELS IN EMBRYONIC MATERNAL DECIDUUM
 CC AND FLOOR PLATE REGION OF THE NEURAL TUBE. LESS IN DEVELOPING
 CC MEMBRANOUS AND ENDOCHONDRAL BONE, SUBMUCOSA OF INTESTINE, DERMIS
 CC OF SKIN AND THE MESENCHYME OF SPLEEN AND LUNG.
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L24755; AAA37306.1; -
 DR HSSP: P00736; IAP0.
 DR MEROPS: M12.005; -
 DR MGD; MGI:88176; Bmp1.
 DR Interpro: IPR001506; Astacin.
 DR Interpro: IPR000152; Asx_hydroxyl.
 DR Interpro: IPR000859; CUB_domain.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR001881; EGF_Ca.
 DR Interpro: IPR000130; Zn_MTPeptidse.
 DR Pfam: PF00008; EGF_2.
 DR Pfam: PF00431; CUB_5.
 DR Pfam: PF01400; Astacin_1.
 DR PRINTS: PR00480; ASTACIN.
 DR SMART: SM00042; CUB_5.
 DR SMART: SM00179; EGF_Ca; 2.
 DR SMART: SM00235; Zinc; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR PROSITE: PS01180; CUB_5.
 DR PROSITE: PS00010; ASX_HYDROXYL; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01187; EGF_Ca; 2.
 DR Growth factor: Cytokine; Repeat; Bone; Cartilage; Hydrolase;
 KM Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;
 KM Glycoprotein; Zymogen.
 FT SIGNAL 1 25
 FT PROPER 26 125
 FT CHAIN 126 991
 FT DOMAIN 126 991
 FT DOMAIN 327 439
 FT DOMAIN 440 551
 FT DOMAIN 552 593
 FT DOMAIN 596 707
 FT DOMAIN 708 748
 FT DOMAIN 752 864
 FT DOMAIN 865 981
 FT METAL 218 218
 FT ACT_SITE 219 219
 FT METAL 222 222
 ZINC (CATALYTIC) (BY SIMILARITY).
 ZINC (CATALYTIC) (BY SIMILARITY).
 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 327 353 BY SIMILARITY.
 FT DISULFID 380 402 BY SIMILARITY.
 FT DISULFID 440 466 BY SIMILARITY.
 FT DISULFID 493 515 BY SIMILARITY.
 FT DISULFID 556 568 BY SIMILARITY.
 FT DISULFID 564 577 BY SIMILARITY.
 FT DISULFID 579 592 BY SIMILARITY.
 FT DISULFID 596 622 BY SIMILARITY.
 FT DISULFID 649 671 BY SIMILARITY.
 FT DISULFID 712 723 BY SIMILARITY.
 FT DISULFID 719 732 BY SIMILARITY.
 FT DISULFID 734 747 BY SIMILARITY.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 991 AA; 11607 MW; 68A1847783A0BB9E CRC64;
 Query Match 11.1%; Score 416.5; DB 1; Length 991;
 Best Local Similarity 27.6%; Pred. No. 1.6e-23;
 Matches 126; Conservative 66; Mismatches 160; Indels 105; Gaps 20;
 QY 11 CGSVATPLGPKRMPDPVEGRGLASPGFPGPEYANDQERRMTLAPPGYRLVFTFHDLELSH 70
 DB 596 CGGFLTKLN-----GSITSPQMPKREYPPNKKNCIQVAPTPYRISLQDFDETEGND 647
 QY 71 LCEIDFVKLSG---AKVLATLGGESTDTERAPGKDTPYSLSSLDITFRSYSEKP 126
 DB 648 VCKYDFEVRGRLNADSKLHGKFCGSEK-----DEVITSOVNNMRYEKSNDYTSK- 698
 QY 127 FTGFEAFYAARDIDCEQVAPCEAPTCDHCHNHILGFGYSCRAGVYLHKKKRC-SALCS 185
 DB 699 -KGFKAHFFS-DKRECSKND---GCQDCVNTGTSGIECCRCRSFVYLDHKKHCKEKGCE 753
 QY 186 GQVFYRSGELSPPEYPPPKLSSCTYSISLEBGFVILDFVESFVETHPETLCPDYF 245
 DB 754 HKV-TSPGTTTSPNMPDKPSKKECTWAISSYTGHHVKLTFVE-MDIESQPE--CAYDH 809
 QY 246 LKIQTDREHNP---FCGKTLPHRIETKSNVYITFVTDESGHPTQWKIHYTAPCP 301
 DB 810 LEVFDGDAKAPVLRGCGSKKRPVLTATGNRMFLREYSDNSVORKFOASHSTEGC--- 866
 QY 302 YPMAPNGHSPVOAKYLLKDSFS-----IFCEGYELLQGLPLKS 343
 DB 867 -----GQVADYKTKTDLYSHAGPCDNNYGVDCENVIYAEGYGV---ELVPQT 913
 QY 344 FTAVCORD-----GSMDRPMACS-IVDCGPPDDLPSGRVEYITGPGVTTYKAVIQ 393
 DB 914 FEVEEFTDCGYDIELFDGYDSTAPRLGRYCGSGSPPEV-----YSAGDSV---LVK 962
 QY 394 YSCETFTYTKVNDGKTYCEADGF---WTSSKSGKSL 427
 DB 963 FHSDDTI-----SKKGFHLRYTSYKFDOTL 987
 RESULT 16
 ID FA9_CANFA STANDARD; PRT; 452 AA.
 AC P19540;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor IX precursor (EC 3.4.21.22) (Christmas factor).
 GN F9.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 TISSUE=Liver;

RX MEDLINE-90311364; PubMed-2367529;
RA Aelred J.H., Read M.S., Brinkhaus K.M., Verma I.M.;
RT "phenotypic correction of factor IX deficiency in skin fibroblasts of
hemophilic dogs.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:5173-5177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-8933338; PubMed-2752110;
RA Evans J.P., Watzke H.H., Ware J.L., Stafford D.W., High K.A.;
RT "Molecular cloning of a cDNA encoding canine factor IX.";
RL Blood 74:207-212(1989).
RN [3]
RP VARIANT HEMOPHILIA B GLU-418.
RX MEDLINE-90099303; PubMed-2481310;
RA Evans J.P., Brinkhaus K.M., Breyer G.D., Reisner H.M., High K.A.;
RT "Canine hemophilia B resulting from a point mutation with unusual
consequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:10095-10099(1989).
CC -1- FUNCTION: FACTOR IX IS A VITAMIN K-DEPENDENT PLASMA PROTEIN THAT
PARTICIPATES IN THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY
CONVERTING FACTOR X TO ITS ACTIVE FORM IN THE PRESENCE OF CA++
IONS, PHOSPHOLIPIDS, AND FACTOR VIIIa.
CC -1- CATALYTIC ACTIVITY: Hydrolyzes one Arg-I-Ile bond in factor X to
form factor Xa.
CC -1- SUBUNIT: FACTOR IX IS ACTIVATED BY FACTOR XIa, WHICH EXCISES THE
ACTIVATION PEPTIDE AND THUS PRODUCES A MOLECULE CONSISTING OF 2
CHAINS HELD TOGETHER BY ONE OR MORE DISULFIDE BONDS.
CC -1- DISEASE: HEMOPHILIA B IS DUE TO DEFECTS IN FACTOR IX.
CC -1- MISCELLANEOUS: CALCIUM BINDS TO THE GAMMA-CARBOXYGLUTAMIC ACID
(GLA) RESIDUES AND, WITH STRONGER AFFINITY, TO ANOTHER SITE,
BEYOND THE GLA DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: M21757; AAA75006.1; -
CC EMBL: M33826; AAA30844.1; -
CC PIR: A30351; A30351.
CC HSSP: P00740; 1CFH.
CC MEROPS: S01.214; -
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR00742; EGF-2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR002383; GLA_blood.
CC InterPro: IPR001254; Ser_protease_Try.
CC InterPro: IPR000294; Vitr_dep_GLA.
CC Pfam: PF00008; EGF_2.
CC Pfam: PF00089; trypsin; 1.
CC Pfam: PF00594; gla; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PRINTS: PR00001; GLABLOOD.
CC SMART: SM00179; EGF_CA; 1.
CC SMART: SM00069; GLA; 1.
CC SMART: SM00020; TYP_SPC; 1.
CC PROSITE: PS00010; ASX_HYDROXIL; 1.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; 2.
CC PROSITE: PS01187; EGF_CA; 1.
CC PROSITE: PS00011; GLU CARBOXYLATION; 1.
CC PROSITE: PS50240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Blood coagulation; Plasma; Serine protease; Calcium-binding;
KW

KW Hydrolase; Glycoprotein; Vitamin K; Gamma-carboxyglutamic acid;
KM Hemophilia; Hydroxylation; Zymogen; Signal; EGF-like domain;
KW Repeat; Disease mutation.
FT SIGNAL 1 21
FT PROPEP 22 39
FT CHAIN 40 183
FT PROPEP 184 217
FT CHAIN 218 452
FT DOMAIN 86 122
FT DOMAIN 123 164
FT DOMAIN 218 452
FT SITE 183 184
FT SITE 217 218
FT MOD_RES 46 46
FT MOD_RES 47 47
FT MOD_RES 54 54
FT MOD_RES 56 56
FT MOD_RES 59 59
FT MOD_RES 60 60
FT MOD_RES 65 65
FT MOD_RES 66 66
FT MOD_RES 69 69
FT MOD_RES 72 72
FT MOD_RES 75 75
FT MOD_RES 79 79
FT MOD_RES 103 103
FT DISULFID 57 62
FT DISULFID 90 101
FT DISULFID 95 110
FT DISULFID 112 121
FT DISULFID 127 138
FT DISULFID 134 148
FT DISULFID 150 163
FT CARBOHYD 197 197
FT CARBOHYD 207 207
FT CARBOHYD 297 297
FT ACT_SITE 258 258
FT ACT_SITE 306 306
FT ACT_SITE 402 402
FT VARIANT 418 418
SQ SEQUENCE 452 AA; 50827 MW; 1F6537C46A960ED CRC64;
Query Match 11.0%; Score 412; DB 1; Length 452;
Best Local Similarity 29.9%; Pred. No 1.3e-23;
Matches 110; Conservative 57; Mismatches 115; Indels 86; Gaps 16;
DB 328 CENGELLOGLHLPLKSFYAVCOQKDSWDRPWACISYDCGPPDDLPSSGRVEYITGPGVTT 387
DB 150 CTTGQLAEDQR-----SCEPAVFPFCGRVS-----VP 177
QY 388 YKAVIQYSCETPTFTMKVNDGKYVCADGFWTSSKGRSLP-VCEPVCGLSARTTGGRIY 446
DB 178 HISMTTRAEFLFSNM-----DYENSTVEKEILDNVQPLNDFP-----RVV 219
QY 447 GGQAKRPGDFPMQVLLILGTTA--AGALLDNVLLTAHVAHYVEQKHDAASLDIRM----- 499
DB 220 GGDARAKGQEPFQVLLNGKVDAPFGGSIINEKVVTAHCTEP-----DKVITIVAG 271
QY 500 --GLKRLSPHYTQAWSEAV--FLHEGYTHDAG-FDNDIALIKLNKVVINSNITPCLP 554
DB 272 EHNTKR---EHTEQKRNVIKTLIHHHSYNTINKYNHDIALLDELDEPLTINSYVTPICIA 328


```

FT DISULFID 714 727 BY SIMILARITY.
FT DISULFID 729 742 BY SIMILARITY.
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 703 730 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 731 986 MISSING (IN ISOFORM BMP1-1).
FT VARSPPLIC 245 302 MISSING (IN ISOFORM BMP1-1).
FT VARSPPLIC 303 986 MISSING (IN ISOFORM BMP1-4).
FT VARSPPLIC 589 622 MISSING (IN ISOFORM BMP1-4).
FT VARSPPLIC 623 986 MISSING (IN ISOFORM BMP1-5).
FT VARSPPLIC 703 717 MISSING (IN ISOFORM BMP1-5).
FT VARSPPLIC 718 986 MISSING (IN ISOFORM BMP1-6).
FT VARSPPLIC 703 823 MISSING (IN ISOFORM BMP1-6).
FT VARSPPLIC 824 986 MISSING (IN ISOFORM BMP1-7).
FT CONFLICT 748 748 D -> N (IN REF. 4).
FT CONFLICT 934 934 R -> S (IN REF. 4).
SQ SEQUENCE 986 AA; 111248 MW; F89201913AC3CBEA CRC64;

Query Match 10.98; Score 411.5; DB 1; Length 986;
Best Local Similarity 28.28; Pred. No. 3.8e-23;
Matches 129; Conservative 65; Mismatches 158; Indels 105; Gaps 22;

11 CGSATPLGPKRPVPGRLASPGFPGYANDQGRMTLTAPGVRLRLVYTHHDELISH 70
591 CGGELTLTLN-----GSTSPGMPKPEYPPNKCIMQVAPYRIISLQDFETESND 642
71 LCEYDFVLLSSG---AKVLTATLGGESTDTERAPGKDTFYSLSLDIRSDYSNEKP 126
643 VCKYDFEVRSGLTADSKLHKGKFCGSEK-----PEVITSQYNNMNVKESDNTYSK- 693
127 FTGEAFYAAEDIDECQVAPGEAPTCDHCHNHLGFCYSCRAGYVLRNKRKTC-SALCS 185
694 -KGFKAHFFS-DKDECSKNDG---GCCODCVNTRGSEYCCGRSGFVLHDKHCKEACD 748
186 GQVTFORSSELSPREYRPRPKLSSCYSTSLBEGFVILDFEVSFVETPELCPIDF 245
749 HKV-TSTSGITTSFNPMDKIPSKKECTWASISPGHRAKLFEME-MIDESQPE--CAADH 804
246 LKTIOTDREHNP---FCGKTLPHRIETKSNVTITFTVDESCHTGMKIHTTAQPCP 301
805 LEVFDGDAKAPVLGRGCGSKKPEPVLATGSRMFLRFYSQNSVQARKGQ---ASHATECG 861
302 YPMAPPMGNHSPVOAKYITLDSFS-----ITCEMGYELLQHLPLKS 343
862 -----GQVRADYKTDLYSHAQFGDNNYPGVDCENVAIAEEGYGV---ELVEQOT 908
344 FTAVACQKOD-----GSMDRPMPACS-IYDCGPPDLPGRVREYITGPGVTYKANIQ 393
909 FEVEBEEDDCGYDIWELDGDIDSTAPRLGRYCGSGPPEEV-----YASAGDSV---LVK 957
394 YSCETFTYMKVNDGXVCEADGF---WTSSKGEKSL 427
958 FHSDDTI-TKK-----GFHLRYTSTFKFODTL 982

```

```

ID ENTK_BOVIN STANDARD; PRT; 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enteroleptinase precursor (EC 3.4.21.9) (Enterokinase).
GN PRESS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kumlantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duoenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RL [2]
RN SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;
RA Lavallie E.R., Rehentulla A., Racie L.A., Dibiase E.A.,
RA Ferenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
subunit of bovine enterokinase."
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RX TISSUE=Intestine;
RL MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
enterokinase."
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
trypsinogen.
CC -1- SUBUNIT: HETERO-DIMER OF A CATALYTIC (LIGHT) CHAIN AND A
MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09859; AAB0026.1; -.
DR EMBL; L19663; AAL16035.1; -.
DR PIR; A61436; A61436.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.156; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000998; MAM_domain.

```

RESULT 18
ENTK_BOVIN

DR	InterPr:	IPR000082:	SEA.domain.
DR	InterPro:	IPR001254:	Scr1.protease.Try.
DR	InterPro:	IPR001190:	Scr1.receptor.
DR	Pfam:	PF00057:	Idl_receptl_a; 2.
DR	Pfam:	PF00089:	trypsin; 1.
DR	Pfam:	PF00431:	CUB; 2.
DR	Pfam:	PF00530:	SRCR; 1.
DR	Pfam:	PF00628:	MAM; 1.
DR	Pfam:	PF01390:	SEA; 1.
DR	PRINTS:	PRO0722:	CHYMOTRYPSIN.
DR	SMART:	SM00042:	CUB; 2.
DR	SMART:	SM00192:	LDla; 2.
DR	SMART:	SM00137:	MAM; 1.
DR	SMART:	SM00200:	SEA; 1.
DR	SMART:	SM00202:	SR; 1.
DR	SMART:	SM00020:	TRYP_SPC; 1.
DR	PROSITE:	PS01180:	CUB; 2.
DR	PROSITE:	PS01209:	LDLRA_1; 2.
DR	PROSITE:	PS50068:	LDLRA_2; 2.
DR	PROSITE:	PS00740:	MAM_1; 1.
DR	PROSITE:	PS50060:	MAM_2; 1.
DR	PROSITE:	PS50024:	SEA; 1.
DR	PROSITE:	PS00420:	SRCR_1; FALSE_NEG.
DR	PROSITE:	PS50287:	SRCR_2; 1.
DR	PROSITE:	PS50240:	TRYPsin.DOM; 1.
DR	PROSITE:	PS00134:	TRYPsin.HIS; 1.
DR	PROSITE:	PS00135:	TRYPsin.SER; 1.
KW	Signal-anchor:		Glycoprotein; Myristate; Hydrolyase;
KV	Serine protease:	Zymogen;	Transmembrane; Repeat; Alternative splicing
FT	CHAIN	1	800
FT	CHAIN	801	1035
FT	DOMAIN	1	18
FT	TRANSMEM	19	47
FT	DOMAIN	48	1035
FT	DOMAIN	54	169
FT	DOMAIN	197	238
FT	DOMAIN	240	350
FT	DOMAIN	358	520
FT	DOMAIN	540	650
FT	DOMAIN	657	695
FT	DOMAIN	694	787
FT	DOMAIN	801	1035
FT	AC1_SITE	841	841
FT	AC1_SITE	892	892
FT	AC1_SITE	987	987
FT	LIPID	2	2
FT	DISULFID	199	212
FT	DISULFID	206	225
FT	DISULFID	219	236
FT	DISULFID	659	671
FT	DISULFID	666	684
FT	DISULFID	678	693
FT	DISULFID	788	912
FT	DISULFID	826	842
FT	DISULFID	926	993
FT	DISULFID	957	972
FT	DISULFID	983	1011
FT	CARBOHYD	116	116
FT	CARBOHYD	147	147
FT	CARBOHYD	170	170
FT	CARBOHYD	194	194
FT	CARBOHYD	233	233
FT	CARBOHYD	263	263
FT	CARBOHYD	264	264
FT	CARBOHYD	404	404
FT	CARBOHYD	456	456
FT	CARBOHYD	486	486
FT	CARBOHYD	519	519
FT	CARBOHYD	550	550
FT	CARBOHYD	646	646
FT	CARBOHYD	698	698
FT	CARBOHYD	722	722

Query Match	10.9%	Score 408.5	DB 1	Length 1035
Best Local Similarity	21.7%	Pred. No. 6.9e-23		
Matches 185	Conservative 111	Mismatches 255	Indels 301	Gaps 38
FT CAROHWD 741 741	N-LINKED (GLCNAC. . .)	(POTENTIAL)		
FT CAROHWD 762 762	N-LINKED (GLCNAC. . .)	(POTENTIAL)		
FT CAROHWD 864 864	N-LINKED (GLCNAC. . .)	(POTENTIAL)		
FT CAROHWD 903 903	N-LINKED (GLCNAC. . .)	(POTENTIAL)		
FT CAROHWD 965 965	N-LINKED (GLCNAC. . .)	(POTENTIAL)		
FT VARSPLIC 166 192	MISSING (IN SHORT ISOFORM)			
FT CONFLICT 808 808	R -> Y (IN REF. 3)			
SEQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;				
Query Match	10.9%	Score 408.5	DB 1	Length 1035
Best Local Similarity	21.7%	Pred. No. 6.9e-23		
Matches 185	Conservative 111	Mismatches 255	Indels 301	Gaps 38
QY 46 RWTLPAPGRIRLRYTFHDELHLCEYDYKLSGAKVATLGGQSDTERA----	100			
DB 270 RMIIRVNOGLSLQLNEDYFN-----TYADVAVINIEGMSKILRASLWSN	315			
QY 101 -GKGTFFVSLGSSLDITPRSDYSNKKPTGPEAFAA-----	136			
DB 316 NPGIIRITS -NOVATATFLIQ-SDSDYIGFVYTTAFNSKELNNEYKINCNEDEGFCFW	372			
QY 137 -EDIDE-----COVAPGAPCPDHHCHHGLGFCYSCRAG-----	170			
DB 373 IODLNDNDNEMETQSTPFPTSGPFPDHFQGN-SGFIYSTPTGGGRERYGLTLPLD	431			
QY 171 -----YVLRHNRKTCALCS-----GOVFQORSGE-195	195			
DB 432 PPEQACLSFWYMGENVYKLSINISSDQNMKTIQKEGNYGQNMVYQVTLNETVEF	491			
QY 196 -----195	195			
DB 492 KVSFGYGNQLISDALDLSLIFYGICNVSVYPEPLVTPPELPTDCGPHDMEPNT	551			
QY 196 -LSPEEYPRPYKLSSTCTYSISLEGESVILDFEVSFVETHPETLCPYD-----FLK	247			
DB 552 TTTSITNFPSIYNQAFCLINLNAQGGKNIQLHFQ-FLDENADVETRDGEGDSLFLA	610			
QY 248 IOTDREHMGPFCKTLPRIETKSNVTVTITFVTDSEGHTGKRIHTS----TAQPCPY	302			
DB 611 VYTG--PCPV-----NDVFSTTRMTVLFTIDMLNAKQGFKAFTTGYGLGIPKCP	660			
QY 303 PMAPRGNHVSPOAKIYIKLDSFTICEGYSIELLOSHLPKSTFVACQ-----KDSMDRP	357			
DB 661 -----EDNFQ--CKDG-----ECIPLVN---LCDFPHCKDSDS--690	690			
QY 358 MPACSLVDCGPDDLPSCGEVYITGPVYTYKAVIQYSCSEFTFYMKYNDGKYVCEADGF	417			
DB 691 -AHCVRLNFGTTDS--SGLVQF-----RDSIMHVAQAEH-WTTOISD--YCOLLGL	737			
QY 418 WTSKGEKSLPVC-----PYCGLSARTTG-----443	443			
DB 738 GP---GNSSVPTFGGGRPYVNLNAPNGSLILTPSQCLLEDSLILLCQNKSKGKLVT	794			
QY 444 -----RIYGGCAKAGDPPMOY-LILGGTTAGA-LLYDNNVLTAAHAYEDCKHDAS---493	493			
DB 795 QEVSEPKYVGGSDSRGCAFWVALKLEDDQOVGASLVRDMLVSAHCVYVGNNMPSKWK	854			
QY 494 -ALDIRMGTLKRLSPHYQAMSEAVEFIEHGYHDGAFNDIALIKLNNKYVINSWITPIC	552			
DB 855 AVLGLHMAS-NLTSQIETRLIDQIVINPHY-NKRRKNDLMMHLEMKVNTTYDIQIC	912			
QY 553 LPRKAESFMRTDIDGTASGAG-LTORGFLEAKNLVYDIPYDHOKCTAAEKEPYPRGS	611			
DB 913 LP--BENQVFPFGRICSLAGMALLIYGSTADVQLEADVPLLSNNKC-----QQQMPPEYN	965			
QY 612 VTANMLCAGLESGGGDCSGDSSGALVFLDSETFEWFYVGGIYVSGSMNGEAGQYVYTK	671			
DB 966 ITENNVACGYEAGYGVDSOGDSGGCLP-MCEENRMWLLAGVTSFG-YOCALPNRPGYVAR	1022			
QY 672 VINIYIPMENTI 683				
DB 1023 VPRTEWIOSFL 1034				

RESULT 19
 ID FA9_HUMAN STANDARD: PRT; 461 AA.
 AC P00740:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor IX precursor (EC 3.4.21.22) (Christmas factor).
 GN F9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=66000558; PubMed=2994716;
 Yoshitake S., Schach B.G., Foster D.C., Davie E.W., Kurachi K.;
 "Nucleotide sequence of the gene for human factor IX (antihemophilic
 factor B).";
 Biochemistry 24:3736-3750(1985).
 [2]
 RP MEDLINE=65190593; PubMed=3857619;
 McGraw R.A., Davis L.M., Noyes C.M., Lundblad R.L., Roberts H.R.,
 RA Graham J.B., Staiford D.W.;
 RT "Evidence for a prevalent dimorphism in the activation peptide of
 RT human coagulation factor IX";
 RT Proc. Natl. Acad. Sci. U.S.A. 82:2847-2851(1985).
 RN [3]
 RP MEDLINE=84236100; PubMed=6329734;
 RA Anson D.S., Choo K.H., Rees D.J.G., Giannelli F., Gould K.G.,
 RA Huddleston J.A., Brownlee G.G.;
 RT "The gene structure of human anti-haemophilic factor IX";
 RT EMBO J. 3:1053-1060(1984).
 RN [4]
 RP MEDLINE=83220788; PubMed=6687940;
 RA Jaye M., de la Salle H., Schamber F., Ballard A., Kohli V.,
 RA Flindell A., Tolstoshev P., Jacocq J.P.;
 RT "Isolation of a human anti-haemophilic factor IX cDNA clone using a
 RT unique 52-base synthetic oligonucleotide probe deduced from the amino
 RT acid sequence of bovine factor IX";
 RT Nucleic Acids Res. 11:2325-2335(1983).
 RN [5]
 RP MEDLINE=36-326 FROM N.A.
 TISSUE=Liver;
 MEDLINE=84300526; PubMed=6089357;
 Jagadeeswaran P., Lavelle D.E., Kaul R., Mohandas T., Warren S.T.;
 "Isolation and characterization of human factor IX cDNA:
 RT identification of Tag I polymorphism and regional assignment.";
 RT Somat. Cell Mol. Genet. 10:465-473(1984).
 RN [6]
 RP MEDLINE=8127096; PubMed=3340835;
 RA Scofield E.S., Koebel D.D., Sarkar G., Sommer S.S.;
 RT "Genomic amplification with transcript sequencing.";
 RT Science 239:491-494(1988).
 RN [7]
 RP MEDLINE=444-461 FROM N.A.
 RA de la Salle C., Charmanlier J.L., Baas M.J., Schwartz A.,
 RA Wiesel M.L., Grunbaum L., Cazenave J.-P.;
 RT "A deletion located in the 3' non translated part of the factor IX
 RT gene responsible for mild haemophilia B";
 RT Thromb. Haemost. 70:370-371(1993).
 RN [8]
 RP MEDLINE=90078229; PubMed=2592373;
 RA Suehiro K., Kawabata S.-I., Miyata T., Takeya H., Takamatsu J.,
 RA Oyata K., Kamiya T., Saito H., Niho Y., Iwanaga S.;
 RT "Blood clotting factor IX BM Nagoya. Substitution of arginine 180 by

RT tryptophan and its activation by alpha-chymotrypsin and rat mast cell
 RT chymase.";
 RT J. Biol. Chem. 264:21257-21265(1989).
 RN [9]
 RP HYDROXYLATION OF ASP-110.
 RX MEDLINE=83308813; PubMed=6688526;
 RA Mcullen B.A., Fujikawa K., Kistel W.;
 RT "The occurrence of beta-hydroxyaspartic acid in the vitamin
 RT K-dependent blood coagulation zymogens.";
 RT Biochem. Biophys. Res. Commun. 115:8-14(1983).
 RN [10]
 RP PROCESSING, ACTIVE SITE, AND CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=78194509; PubMed=659613;
 RA di Scipio R.G., Kurachi K., Davie E.W.;
 RT "Activation of human factor IX (Christmas factor).";
 RT J. Clin. Invest. 61:1528-1538(1978).
 RN [11]
 RP CALCIUM-BINDING DATA.
 RX MEDLINE=84185715; PubMed=6425296;
 RA Morita T., Isaacs B.S., Eason C.T., Johnson A.E.;
 RT "Derivatives of blood coagulation factor IX contain a high affinity
 RT Ca²⁺-binding site that lacks gamma-carboxyglutamic acid.";
 RT J. Biol. Chem. 259:5698-5704(1984).
 RN [12]
 RP ERRATUM.
 RA Morita T., Isaacs B.S., Eason C.T., Johnson A.E.;
 RT J. Biol. Chem. 260:2583-2583(1985).
 RN [13]
 RP SIGNAL SEQUENCE CLEAVAGE SITE.
 RX MEDLINE=86189947; PubMed=609023;
 RA Bentley A.K., Rees D.J.G., Rizza C., Brownlee G.G.;
 RT "Defective propeptide processing of blood clotting factor IX caused
 RT by mutation of arginine to glutamine at position -4.";
 RT Cell 45:443-448(1986).
 RN [14]
 RP STRUCTURE OF CARBOHYDRATE ON SER-99.
 RX MEDLINE=90062160; PubMed=2511201;
 RA Nishimura H., Kawabata S., Kistel W., Hase S., Ikenaka T., Takao T.,
 RA Shimonishi Y., Iwanaga S.;
 RT "Identification of a disaccharide (Xyl-Glc) and a trisaccharide
 RT (Xyl2-Glc) O-glycosidically linked to a serine residue in the first
 RT epidermal growth factor-like domain of human factors VII and IX and
 RT protein Z and bovine protein Z.";
 RT J. Biol. Chem. 264:20320-20325(1989).
 RN [15]
 RP STRUCTURE OF CARBOHYDRATE ON SER-99.
 RX MEDLINE=91344709; PubMed=2129367;
 RA Iwanaga S., Nishimura H., Kawabata S., Kistel W., Hase S., Ikenaka T.;
 RT "A new trisaccharide sugar chain linked to a serine residue in the
 RT first EGF-like domain of clotting factors VII and IX and protein Z.";
 RT Adv. Exp. Med. Biol. 281:121-131(1990).
 RN [16]
 RP STRUCTURE OF CARBOHYDRATE ON SER-107.
 RX MEDLINE=92388094; PubMed=1517205;
 RA Nishimura H., Takao T., Hase S., Shimonishi Y., Iwanaga S.;
 RT "Human factor IX has a tetrasaccharide O-glycosidically linked to
 RT serine 61 through the fucose residue.";
 RT J. Biol. Chem. 267:17520-17525(1992).
 RN [17]
 RP PHOSPHORYLATION OF SER-114.
 RA Harris R.J., Papac D.I., Truong L., Smith K.J.;
 RT "Partial phosphorylation of serine-68 in EGF-1 of human factor IX.";
 RT (in) Abstracts of Xth International conference on methods in protein
 RT structure analysis, pp.50-50, Amnecy (1996).
 RN [18]
 RP POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE=20575397; PubMed=11133752;
 RA Arruda V.R., Hagstrom J.N., Deitch J., Helman-Patterson T.,
 RA Camire R.M., Chu K., Fields P.A., Herzog R.W., Couto L.B.,
 RA Larson P.J., High K.A.;
 RT "Posttranslational modifications of recombinant myotube-synthesized
 RT human factor IX.";
 RT Blood 97:130-138(2001).

```

RN [19]
RP STRUCTURE BY NMR OF 47-93.
RX MEDLINE=95229607; PubMed=7713897;
RA Freedman S.J., Furie B.C., Furie B., Baleja J.D.;
RT "Structure of the metal-free gamma-carboxyglutamic acid-rich membrane
binding region of factor IX by two-dimensional NMR spectroscopy.";
RL J. Biol. Chem. 270:7980-7987(1995).
RN [20]
RP STRUCTURE BY NMR OF 47-93.
RX MEDLINE=96032604; PubMed=7547952;
RA Freedman S.J., Furie B.C., Furie B., Baleja J.D.;
RT "Structure of the calcium ion-bound gamma-carboxyglutamic acid-rich
domain of factor IX.";
RL Biochemistry 34:12126-12137(1995).
RN [21]
RP STRUCTURE BY NMR OF 47-93.
RX MEDLINE=96279169; PubMed=8663165;
RA Freedman S.J., Blostein M.D., Baleja J.D., Jacobs M., Furie B.C.,
Furie B.;
RT "Identification of the phospholipid binding site in the vitamin K-
dependent blood coagulation protein factor IX.";
RL J. Biol. Chem. 271:16227-16236(1996).
RN [22]
RP STRUCTURE BY NMR OF 47-93.
RX MEDLINE=97199336; PubMed=9047312;
RA Li L., Darden T.A., Freedman S.J., Furie B.C., Furie B., Baleja J.D.,
Smith H., Hskey R.G., Pedersen L.G.;
RT "Refinement of the NMR solution structure of the
gamma-carboxyglutamic acid domain of coagulation factor IX using
molecular dynamics simulation with initial Ca2+ positions determined
by a genetic algorithm.";
RL Biochemistry 36:2132-2138(1997).
RN [23]
RP STRUCTURE BY NMR OF 91-133.
RX MEDLINE=91308127; PubMed=1854745;
RA Huang L.H., Cheng H., Pardi A., Tam J.P., Sweeney W.V.;
RT "Sequence-specific 1H NMR assignments, secondary structure, and
location of the calcium binding site in the first epidermal growth
factor like domain of blood coagulation factor IX.";
RL Biochemistry 30:7402-7409(1991).
RN [24]
RP STRUCTURE BY NMR OF 92-130.
RX MEDLINE=93284090; PubMed=1304885;
RA Baron M., Norman D.G., Harvey T.S., Handford P.A., Mayhew M.,
Tse A.G.D., Brownlee G.G., Campbell I.D.C.;
RT "The three-dimensional structure of the first EGF-like module of
human factor IX: comparison with EGF and TGF-alpha.";
RL Protein Sci. 1:81-90(1992).
RN [25]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 92-130.
MEDLINE=95330802; PubMed=7606779;
RA Rao Z., Handford P., Mayhew M., Knott V., Brownlee G.G., Stuart D.;
RT "The structure of a Ca(2+)-binding epidermal growth factor-like
domain: its role in protein-protein interactions.";
RL Cell 82:131-141(1995).
RN [26]
RP MOLECULAR PATHOLOGY OF HEMOPHILIA B.

```

```

QY 512 AMSEAVFIHEGYTHDAEPNDIALIKLNKRVINSNTPTICLPKEAESFMRDIDGTAS 571
DB 305 -YNAAI--NKYNH-----DIALEDEPLVNSVTPICADKEKTNFLRFGSDYVS 354
QY 572 GWG-LTORGFRLARMLVDPIDYDHOKTAAYEKPPYPRCSVTANMLCAGLESQKSCSR 630
DB 355 GWGRVPHKGRSALVLOYLRPLVDNRATCLRS-----TKRTIYNMNCAGFHEGGRSCQ 408
QY 631 GDSCGALVPLDSETE-RWEVGGIVSGSMNCGEAGQYVTKYINIPWIE 680
DB 409 GDSCGPHV---TEVEGTSFLTGISMGE-ECAMKGKGIYTKVSRVYMWIK 455

RESULT 20
THRB_MOUSE STANDARD: PRT: 618 AA.
ID THRB_MOUSE
AC P19221;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Prothrombin precursor (EC 3.4.21.5).
GN F2 OR CF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=91025551; PubMed=2222810;
RA Friener Degen S.J., Schaffer L.A., Jamison C.S., Grant S.G.,
Fitzgibbon J.J., Pai J.-A., Chapman V.M., Elliott R.W.;
RT "Characterization of the cDNA coding for mouse prothrombin and
localization of the gene on mouse chromosome 2.";
RL DNA Cell Biol. 9:487-498(1990).
RN [2]
RP SEQUENCE OF 384-618 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92212913; PubMed=1557383;
RA Banfield D.K., Macgillivray R.T.;
RT "Partial characterization of vertebrate prothrombin cDNAs:
RT amplification and sequence analysis of the B chain of thrombin from
nine different species.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
CC -1- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Gly: activates
CC fibrinogen to fibrin and releases fibrinopeptide A and B.
CC -1- PPM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC CHARGED PHOSPHOLIPID SURFACE, WHICH IS ESSENTIAL FOR THE CONVERSION
CC OF PROTHROMBIN TO THROMBIN.
CC -1- MISCELLANEOUS: PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A
CC PHOSPHOLIPID MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN &
CC FACTORS VA & XA IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES
CC THE ACTIVATION PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT &
CC HEAVY CHAINS. THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR
CC V ITSELF HAS TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF
CC THROMBIN.
CC -1- MISCELLANEOUS: THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL
CC FRAGMENT (FRAGMENT 1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION
CC BY FACTOR XA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

FT	DISULFID	561	591	BY SIMILARITY
FT	CARBOHYD	122	122	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	144	144	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	413	413	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. . .)
SO	SEQUENCE	618 AA;	70268 MW;	B89F719A9AFD01ED CRC64;
Query Match 10.8%; Score 405; DB 1; Length 618;				
Best Local Similarity 25.6%; Pred. No. 6.5e-23;				
Matches 163; Conservative 72; Mismatches 232; Indels 170; Caps 29;				
OY	131 EAFYAEIDIDECQVAPGAPPCDH-----HCNHHUGFPCSCRCAGYVLRHN	176		
Db	70 EAFELIESPDQTDVFWAATYVCDSYRKRPREFPMDCLEGRCAMDJLVNPL-----	118		
OY	177 KRTGALCSGOVGFQNRSG---ELSSPEYPRPYPKLSCTGYSISLEGRSVILDFVESFDV	233		
Db	119 -----GIVVNTHTHTIQCOLMRSRPHK-PEINSTTHPGA-----DLKENCR	159		
OY	234 ETHPEPLCPDYFLKIQTDREBH--GPFCKGKTLPHRIETKSNVTYITFEVTDSCGHTGKI	291		
Db	160 NPDSSTGTGPMCTYTPDPTVIRECSYVPCG-----QGRTTYVMTTPRSGSKDN-----	207		
OY	292 HTSTAQCPPIPMAPRPNCHSVYQAKYTLKDSFSIFCTGTELLQGLPLKSFPAVC---	348		
Db	208 -----LSPPILG-----QCLTFERG--RLYOGNLAVTTLGSPCLPW	239		
OY	349 -----QKDSMPRMPACSIYD--CGPPDDLPSCGVEVYITG--PGVTTYKAVIYQSC	397		
Db	240 NSLPAAKTLSTKTDQDF---PEVKLVENFCRPNMDDEGAMCYVAGQPDPEY-CNLVNT-CE	294		
OY	398 ETFYMKVNDGKYVCE-----ADGEWTSKGEKSLVCEPCVL-----SARTT--	441		
Db	295 EA-----VGEENVYDDESIAGRITDAEHHTFENETPFLGELADGCLRLFEKSLAKDTTE	349		
OY	442 -----GRIYGGQAKPDGFPWOVILIGT---TAAGALLYDNVLTAAHAV----	485		
Db	350 KELDSYIDGRIVEGMDAEKGIAPQVALLFRKSPQELLGASLISDRWVLTAAHCILYPP	409		
OY	486 YEOKHDASALDIRMGTLRKLSPHITQ-----AMSAVYTHGCTYHDAGFDDIDIALIKLN	540		
Db	410 WDKNETENDLLVRIG--KHSRTRYERNNEKISMELKIVVHPRYMWRREMIJDRDIALLK	467		
OY	541 KVIINSNTPICLPRKE--AESPMRTDDIGTASGMLTQRGFLAR-----NLWYVDIPI	592		
Db	468 PVPESDTHPVCPLPKQCVYVTSILRLGKYGKRAYIGWGNLEETVTNNINLQPSYLOVYNLPI	527		
OY	593 VDHOKCTAAEKEPPYPRGCSVTANMLACGL---ESGKSDSCRGSDGALVFLDSETERMFV	649		
Db	528 VERPYCKRAS-----TRIRIDNMCCAGFYKNDTRKGDACGSDSGSPVNMKSPFNRRYQ	581		
OY	650 GGIYVSGSMNCGEAGQYGVYTKVINYIWIENIISDF	686		
Db	582 MGIVSWGEGCDRKGKGYFTYHVFRLKRIQKVIDOF	617		
RESULT 21				
HPT_CANFA				
ID	HPT_CANFA	STANDARD:	PRT:	329 AA.
AC	PI9006:			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	01-OCT-1986 (Rel. 34, Last annotation update)			
DE	Haptoglobin alpha and beta chains.			
GN	HP.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	SEQUENCE.			
RA	MEDLINE=93054403; PubMed=1429498;			
RA	Kumazaki T., Urushibara N., Ishii S.-I.;			

```

"Antio acid sequence and disulfide-bridge location of canine
haptoglobin.";
RL J. Biochem. 112:11-19(1992).
[2]
RN SEQUENCE OF 85-124.
RX MEDLINE=77025019; PubMed=975782;
RA Kurosky A., Kim H.H., Touchstone B.;
RT "Comparative sequence analysis of the N-terminal region of rat,
rabbit, and dog haptoglobin beta-chains.";
RL Comp. Biochem. Physiol. 55B:453-459(1976).
CC -1- FUNCTION: HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN,
PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
ACCESSIBLE TO DEGRADATIVE ENZYMES.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Synthesized in the liver, then released
in plasma.
CC -1- MISCELLANEOUS: THE HEMOGLOBIN BINDING SITE IS LOCATED IN THE BETA
CHAIN.
CC -1- SIMILARITY: THE BETA CHAIN IS CLEARLY RELATED TO SERINE PROTEASES,
BUT HAPTOGLOBIN HAS NO PROTEOLYTIC ACTIVITY. PROBABLY BECAUSE THE
POSITIONS CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES IN
PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPTOGLOBIN.
CC PIR: JX0223; JX0223.
DR PIR: JX0224; JX0224.
DR PIR: B26503; B26503.
DR HSSP: P00734; 1B7X.
DR MEROPS: S01.972; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protase_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_Sec; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Glycoprotein; Serine protease homolog; Plasma; Hemoglobin-binding;
KW Layer.
FT CHAIN 1 83 HAPTOGLOBIN ALPHA CHAIN.
FT PROPEP 84 84 REMOVED (IN AT LEAST 80% OF THE ALPHA
CHAINS).
FT CHAIN 85 329 HAPTOGLOBIN BETA CHAIN.
FT DOMAIN 85 329 SERINE PROTEASE.
FT DISULFD 34 68 INTERCHAIN (ALPHA-BETA).
FT DISULFD 72 189
FT DISULFD 232 263
FT DISULFD 274 304
FT CARBOHYD 9 9 N-LINKED (GLCNAC. . .).
FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .).
SEQUENCE 329 AA; 36457 MW; 86E32CA3E9CC2C48 CRC64;
Query Match 10.7%; Score 403; DB 1; Length 329;
Best Local Similarity 31.3%; Pred. No. 4,2e-23;
Matches 104; Conservative 45; Mismatches 145; Indels 38; Gaps 11;
QY 369 PDDPSGRVEYITGCVTYTKAVIOYSCETTYTKAVNDKVCYCEADGFWTSSKKEKSLP 428
D 18 PVLVEIGYVEH-----MIRYQCKPFYKLTHTGGVYTLNSEKHWTKKAVGEKLP 66
QY 429 VCEPVCGLSARTTG--GRITGGQKAKPDDEPQVLLIG--GTTAGALLIDYWWVLTAAHA 484
D 67 ECEAVCGKPKNEVDVOYHIMGSDVDAKSPFOAKMVSNNHTSATLLINQEWLLTAKN 126
QY 485 VY-EQKHASALDIMGTKRLSPHYTQWSEAVYIHGGYTHDAQFDDNDIALIKNNKYV 543
D 127 LFLGKDKDAKANDI-APLKLKLVYKGNQLVEYKVLVLDHDS-----KYDIGIKLKQKVP 180
QY 544 INSNITPICLPKKEAESMTRTDIDGTASGGLITQGFARLNMLVYDIPVHQKTAAYE 603
D 181 IDERMPALCPSKD--YAEVGRIGTVSGWGRNSNFNTLLKLYMPLVADQDKCVQHYE 237
QY 604 -----KPRYPRG---SVTANMLCAGLESGKSCGDSGGLVFLDSETERFVVGCI 652

```

Db	238	GSTVEKKSPPSPVGVOPILNHEHFCAGMSKFFQEDTCYGDAGSAFAVHDDDEDTFWAAGI	297
Oy	653	VSMGSMCGEAGOTGVTKVNTNTPWENTIS	684
Db	298	LSF-DKSCCTVA-EYGVYKVPVSLAWOETIA	327
RESULT	22		
FA10_HUMAN	STANDARD;	PRT;	488 AA.
AC	P00742; Q14340;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	01-OCT-1989 (Rel. 12, Last sequence update)		
DE	Coagulation factor X precursor (EC 3.4.21.6) (Stuart factor).		
GN	F10.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91216473; Pubmed=1902434;		
RA	Messier T.L., Pittman D.D., Long G.L., Kaufman R.J., Church W.R.;		
RT	"Cloning and expression in COS-1 cells of a full-length cDNA encoding		
RL	human coagulation factor X.";		
RL	Gene 99:291-294(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=87026600; Pubmed=3768336;		
RA	Leytus S.P., Foster D.C., Kurachi K., Davie E.W.;		
RT	"Gene for human factor X: a blood coagulation factor whose gene		
RT	organization is essentially identical with that of factor IX and		
RL	protein C.";		
RL	Biochemistry 25:5098-5102(1986).		
RN	[3]		
RP	SEQUENCE OF 13-488 FROM N.A.		
RX	MEDLINE=85216545; Pubmed=2582420;		
RA	Fung W.R., Hay C.W., McGilivray R.T.A.;		
RT	"Characterization of an almost full-length cDNA coding for human		
RT	blood coagulation factor X.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 82:3591-3595(1985).		
RN	[4]		
RP	SEQUENCE OF 19-488 FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=86221713; Pubmed=3011603;		
RA	Kaul R.K., Hildebrand B., Roberts S., Jagadeeswaran P.;		
RT	"Isolation and characterization of human blood coagulation factor X		
RL	cDNA.";		
RL	Gene 41:311-314(1986).		
RN	[5]		
RP	SEQUENCE OF 41-179.		
RX	MEDLINE=83257207; Pubmed=6871167;		
RA	Mcullien B.A., Fujikawa K., Kisiel W., Sasagawa T., Howard W.N.,		
RA	Kwa E.Y., Weinstein B.;		
RT	"Complete amino acid sequence of the light chain of human blood		
RT	coagulation factor X: evidence for identification of residue 63 as		
RL	beta-hydroxyaspartic acid.";		
RL	Biochemistry 22:2875-2884(1983).		
RN	[6]		
RP	SEQUENCE OF 115-488 FROM N.A., AND TISSUE SPECIFICITY.		
RC	TISSUE=Liver;		
RX	MEDLINE=84222026; Pubmed=6587384;		
RA	Leytus S.P., Chung D.W., Kisiel W., Kurachi K., Davie E.W.;		
RT	"Characterization of a cDNA coding for human factor X.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 81:3699-3702(1984).		
RN	[7]		
RP	SEQUENCE OF 183-234, AND CARBOHYDRATE-LINKAGE SITES.		
RX	MEDLINE=94062825; Pubmed=8243461;		
RA	Inoue K., Morita T.;		
RT	"Identification of O-linked oligosaccharide chains in the activation		
RT	peptides of blood coagulation factor X. The role of the carbohydrate		
RT	moieties in the activation of factor X.";		


```

Cc 95 --CONOGCKRGLGEYTCLEGF-----EGKNC-----ELFTRKLSLDN-----133
Cc 208 LSSCTYSILEEGSVLLDFVESFDVETHPTLPYDLKIQTRERHGP-CCK-TLPH 265
Cc 134 -GDGD-OFCEBONSVCSCARGTTLADNGKACIP-----TGPPCKQTLER 179
Cc 266 RIETKSNVTITFTVDESQHTGKIRHYTSTAQPCPYMPAPNGHVSFQAKYLLKDSFS 325
Cc 180 R---KRYAQAATSSSGEAPDSITWK-----PYDAA---DLDPTEPFLLD---219
Cc 326 IFCETPELLQGLPLKSFYAVCGKDSWDRPMACSIIVDCGPPDDLPSCREVEYITGPV 385
Cc 220 -FNQTOPE-----226
Cc 386 TTYAAVYQYSCETFTYMKVNDGKYVCEADGFTSSGKESLPVCEPVCGLSARTTGRI 445
Cc 227 -----RDNNLT-----RI 235
Cc 446 YGGGAKRGPDPQVLLIGTT---AAGALLYDNMVLTAHAVYEOKHDAALDIRMGTL 502
Cc 236 VGGGECCKDCEPQWALLINENEGCGTITSEFYILTAHCLYQAKR---FKVYVGR 291
Cc 503 KRISPHYTOAMS--EAVFIHEGYTHDAGFNDIALIKLNKRVINSNITPICLPKKE-AE 559
Cc 292 NTEQEGGEAAVHEVEVYIKHNRFTKET-YDEDIAYLRLKPTIFRMANVAPACLPEDMAE 350
Cc 560 SFMTDIDIGTASGGLT-QRGFLARNLMKYVDIPYDHOKCTAAAEKPPYPGSTANMLC 618
Cc 351 STLTKQTKGTIVSGFRHEKROSTRLKMLEVPYVNSCLSL-----SSEITQWMC 404
Cc 619 AGLESCKDSCGSGALVFLDSETERMEFGIVSGMSMCGAGGYVTKYNTIPW 678
Cc 405 AGYTKQEDACQSGSGPHYTRFKDT--YFVTVGIVSMGE-CCAKKKGITTKYATLKW 461
Cc 679 IE 680
Cc 462 ID 463

```

RESULT 23

FA9_CAVPO STANDARD; PRT; 285 AA.

AC p16295;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DT Coagulation factor IX (EC 3.4.21.22) (Christmas factor) (Fragment).

CC P9.

CC *Canis familiaris* (Guinea pig).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.

CC NCBI_TaxID=10141;

CC [1]

CC SEQUENCE FROM N.A.

CC MEDLINE=90152675; PubMed=2303254;

CC Sakar G., Koebel D.D., Sommer S.S.;

CC "Direct sequencing of the activation peptide and the catalytic domain

CC of the factor IX gene in six species."

CC Genomus 6:133-143(1990).

CC -FUNCTION: FACTOR IX IS A VITAMIN K-DEPENDENT PLASMA PROTEIN THAT

CC PARTICIPATES IN THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY

CC CONVERTING FACTOR X TO ITS ACTIVE FORM IN THE PRESENCE OF CA++

CC IONS, PHOSPHOLIPIDS, AND FACTOR VIII.

CC -CATALYTIC ACTIVITY: Hydrolyzes one Arg-Ile bond in factor X to

CC form factor Xa.

CC -SUBUNIT: FACTOR IX IS ACTIVATED BY FACTOR XIa, WHICH EXCISES THE

CC ACTIVATION PEPTIDE AND THUS PRODUCES A MOLECULE CONSISTING OF 2

CC CHAINS HELD TOGETHER BY ONE OR MORE DISULFIDE BONDS.

CC -SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb.ch).

CC EMBL; M26237; AAA37037.1; -

CC HSSP; p16293; 1PFx.

CC MEROPS; S01.214; -

CC InterPro; IPR001254; Ser_protease_Try.

CC InterPro; IPR000294; VitK_dep_GLA.

CC Pfam; PF00089; trypsin; 1.

CC SMART; SM00020; Tryp_spc; 1.

CC PROSITE; PS00011; GLU_CARBOXYLATION; PARTIAL.

CC PROSITE; PS00240; TRYPsin_DOM; 1.

CC PROSITE; PS00134; TRYPsin_HIS; 1.

CC DR PROSITE; PS00135; TRYPsin_SER; 1.

CC DR Blood coagulation; Plasma; Serine protease; Calcium-binding;

CC KW Hydrolyase; Glycoprotein.

CC FT NON_TER 1 1

CC FT ACT_SITE 99 99

CC FT ACT_SITE 147 147

CC FT ACT_SITE 243 243

CC FT CARBOHYD 25 25

CC FT CARBOHYD 45 45

CC FT CARBOHYD 127 127

CC FT CARBOHYD 138 138

CC FT NON_TER 285 285

CC SQ SEQUENCE 285 AA; 32032 MW; 6836DBD317C11C60 CRC64;

Query Match 10.7%; Score 402.5; DB 1; Length 285;

Best Local Similarity 32.3%; Pred. No. 3,8e-23;

Matches 104; Conservative 58; Mismatches 105; Indels 55; Gaps 15;

Cc 372 LPSGRVETITGPVYTKAVYQYSCETFTYMKVNDGKYVCEADGFTSSGKESLPVCE 431

Cc 4 IPVSKEHNRNANAFISRMGYVNFDETDIWDNDDD-----ETIWDNST-ESTKPSD 55

Cc 432 PYGSLARTTGGRTYGGOKRKPDPQVLLIGTTA--AGALLYDNMVLTAHA-----484

Cc 56 FF-----RVGSEDAKPGQFPQVLLNGETEAFCGSGSTVNEKWTAAHCILPGI 105

Cc 485 ---VEQKHDAALDIRMGTLLKRLSPHYTOAMSFAVFIHEGYTHDAGF--NDIALIKL 538

Cc 106 KIEVACKHN---IEKEDDEQR--RMVQ-----ILHHSY--NMSFNKSYDIALLEL 153

Cc 539 NKKVYINSNITPICLPKKEASEFMRDIDGTASGAG-LTQRGFLARNLMKYVDIPYDHOK 597

Cc 154 DKPLSLNSVYTPICIANREYTNIFLKFGAGYVSGWGLSQGRTASTILOYLRLVPLVDRAT 213

Cc 598 CTAAYEKPPRPGSVTANMCAGLESGKDSGSGALVFLDSETERF-VFGGIYSWG 656

Cc 214 CLRS-----TKFTIYNMKCAGHSGRDSGSGPHV---TEVEGTNPLTGLISWG 264

Cc 657 SMNCGEAGGYGYVTKYNTIPW 678

Cc 265 E-DEAMKKGKGYITKYSRYVNW 285

RESULT 24

HPT_ATEGE STANDARD; PRT; 347 AA.

AC p50417;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Haptoglobin precursor.

GN HP.

OS *Ateles geoffroyi* (Black-handed spider monkey).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.

CC NCBI_TaxID=9509;

CC [1]

RP SEQUENCE FROM N.A.
 RA Erickson L.M., Becker L., Maeda N.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN
 CC PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
 CC KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
 CC ACCESSIBLE TO DEGRADATIVE ENZYMES (BY SIMILARITY).
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: Synthesized in the liver, then released
 CC IN PLASMA (BY SIMILARITY).
 CC -1- SIMILARITY: ALTHOUGH THE BETA CHAIN IS CLEARLY RELATED TO SERINE
 CC PROTEASES, HAPTOGLOBIN HAS NO ENZYMATIC ACTIVITY. POSITIONS
 CC CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES OF THE
 CC PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPTOGLOBIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U04852; AAA03727.1; -
 CC HSSP: P08709; 1PAK.
 CC -----
 DR HEMOPS: S01.972; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Ser-protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00084; sush1.1.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP.1
 DR SMART: SM00020; Tryp_Spc.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR Glycoprotein: Serine protease homology; Plasma; Hemoglobin-binding;
 DR Acute phase; Liver; Signal.
 DR SIGNAL: 1 18
 FT CHAIN 19 347
 FT CHAIN 19 101
 FT CHAIN 103 347
 FT DOMAIN 103 347
 FT DISULFID 33 33
 FT DISULFID 52 86
 FT DISULFID 90 207
 FT DISULFID 250 281
 FT DISULFID 292 322
 FT CARBOHYD 125 125
 FT CARBOHYD 148 148
 FT CARBOHYD 152 152
 FT CARBOHYD 182 182
 SQ SEQUENCE 347 AA: 38476 MW: 208480967C17C21F CRC64;
 Query Match 10.7%; Score 402; DB 1; Length 347;
 Best Local Similarity 28.1%; Pred. No. 5,3e-23;
 Matches 98; Conservative 55; Mismatches 132; Indels 64; Gaps 12;

DB 187 ----DIGLIKLDKVPNERVMPICLPSKD---YAEVGRVGVSGMGRNANFKFTDLKY 239
 QY 588 VDIPIVHOKCTAYE-----KPPYRGS---YANMLCGLSGGDSGDSGGA 636
 DB 240 VMLPVAQYQCVKHKEGYSTYPEKTPKSPVGGQDILNEHFFCAQMSXYQEDTCYGDGSA 299
 QY 637 LVPLDSETERFVFGVGVSGMNCGEAGQGVYKRVINYIPMIENISD 685
 DB 300 FAVHDLDEDTMYAGLLISF-DKSGCA-EKGVYKATSIDMWOKTIAE 346
 RESULT 25
 HEPES_RAT STANDARD; PRT; 416 AA.
 ID HEPES_RAT
 AC 005511;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Serine protease hepsin (EC 3.4.21.-).
 GN HPN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=93305733; PubMed=8318546;
 RA Farley D., Raymond F., Nick H.;
 RT "Cloning and sequence analysis of rat hepsin, a cell surface serine
 RL proteinase";
 RL Biochim. Biophys. Acta 1173:350-352(1993).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 CC OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X70900; CA50256.1; -
 CC PIR: S32013; S32013.
 DR PIR: S33777; S33777.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.224; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Ser-protease_Try.
 DR InterPro: IPR001190; Strc_receptor.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; Tryp_Spc.1.
 DR SMART: SM00020; Tryp_Spc.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR PROSITE: PS00134; TRYPSIN_HIS.1.
 DR PROSITE: PS00135; TRYPSIN_SER.1.
 DR Hydrolyase: Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 161
 FT CHAIN 162 416
 FT DOMAIN 1 16
 FT TRANSREM 17 43
 FT DOMAIN 44 416
 FT DOMAIN 162 416
 FT ACT_SITE 202 202
 FT ACT_SITE 256 256
 FT ACT_SITE 352 352
 FT DISULFID 152 276
 FT DISULFID 187 203
 BY SIMILARITY.

FT	DISULEFID	321	337	BY SIMILARITY.
FT	DISULEFID	348	380	BY SIMILARITY.
FT	CARBOHYD	111	111	N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ	SEQUENCE	416 AA;	44926 MW;	ESA9F8FA9550E180 CRC64;

Query Match	10.7%	Score 401.5	DB 1	Length 416
Best Local Similarity	30.1%	Pred. No. 7.3e-23		
Matches 123, Conservative	50	Mismatches 133	Indels 102	Gaps 18

```

QY 311 VSPQAKIILND-----SFSJFC-----ETWELLOGHPLPLSFSA----- 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 LSPBDSRLIVADTKKEGTRMRLLCSSRSNARAAGLCCEMGFLRALHSELDVRTAGANGTS 114

QY 347 --VCQKDGSW--DRPMACSIYDCGPDDLPSSGRVHEYITGPVYTKKAVIYXSCFEFFY 401
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 GFECVDEEGLPLAQRLDLDVISVCCD-----PRGR-FLTA-----TCQD--- 151

QY 402 TMKYNVDGKYVCEADGFWTSSKGEKSLPYCEPVCGLSARTTGKRGYQKAKKPDGFPMOVL 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      152 -----C-----GRKKRPV-----DRIVGQDSSLRWPMQVS 178

462 IL--GGTTAAGALLYDNMVMVLTAAAYVEOKHDASALDIRMGTLRKLSPIHTQAMSEVFI 519
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 LRYDGTILHCGSSLGSMVLTAAACPFEDRNVRNRSRMWRFVGAARVSPHAAVLQGVAVIY 238

QY 520 HEGYT--HDAGFD--NDIALIKLNKNVIVINSNTPICLPRKEASEFMRDDIGTASGWG 574
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 HGGYLRPFPRDPTIDENSNDIALVHLSSSLPLEYIQRPCLP--AAGQALVDGKCYCTIYGMG 266

QY 575 LTO--RGFLARLMLVVDPIYDHOKCTAYEKRPYPRGCVTANMLCAGLESGKDGSCRDGS 633
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 NTQYGGQAAVVLQEARVPIISNEVCNS---PDYGNQIIPKMECAVYPEGIGDADCGDS 352

QY 634 GGALVFLD--SETERMEVGVSGWMSNCCBAGQYGYVTVYVINTIPMI 679
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 GGHHVCECDRIISGTRMKRLCGIVSNGT--GCALAKRPGVYTVYVIDPREVI 399

```

Search completed: January 23, 2003, 22:18:12
Job time : 45 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 23, 2003, 22:20:36 ; Search time 326 Seconds
(without alignments)
4738.868 Million cell updates/sec

Title: US-09-874-198-2
Perfect score: 3758
Sequence: 1 MRLTLGLLCSVATPLGP.....GYTRVINYIPWNIISDF 686

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Search: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 75 summaries

Command line parameters:
-MODEL=framer.p2n.model -DEV=tlp
-Q=/cgn2.1/USPTO.spool/US09874196/runat.22012003_085000_10841/app.query.fasta.1.839
-DB=N.Geneseq.101002 -OFMT=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=75 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=75
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09874198.ecgn1.1.79.@runat.22012003_085000_10841 -NCPU=6 -ICPU=3
-NO_XIPXY -NO_MAP -LARGEFUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	3738.5	99.5	2472	24	AAD24224	Human MASP-2 CDNA
2	1617	43.0	3106	23	AAS85321	Human encoding novel
3	1608	42.8	3863	22	AAH43312	CDNA encoding prot
4	1600	42.6	3895	22	AAF83985	Human serine prote
5	1412	37.6	2492	20	AAK24297	Human complement c
6	1412	37.6	2493	22	ABA83114	Human complement c
7	1412	37.6	2493	24	ABN95758	Gene #2256 used to
8	1238	32.9	2647	24	ABN95607	Gene #2105 used to
9	1238	32.9	2725	21	AACT7788	Human cancer assoc
10	1214	32.3	2908	24	ABK63441	Rat sequence diffe
11	1086.5	28.9	2358	23	AAS92433	DNA encoding novel
12	873	23.2	2063	21	AAE18150	Lung cancer associ
13	689.5	18.3	1148	21	AAE18149	Lung cancer associ
14	536.5	14.3	371	22	AAK53540	Murine transport a
15	498	13.3	3094	21	AAA93124	Human secreted pro
16	466	13.2	1667	22	AAK14091	Human FCTR5A DNA s
17	436	13.2	1691	22	AAK14092	Human FCTR5B DNA s
18	430	13.0	3147	20	AAK87815	Tumour antigen der
19	430	13.0	3147	22	AAK13113	Human membrane-ty
20	430	13.0	3147	22	AAK13155	Human membrane-ty
21	430	13.0	3147	22	AAH23601	Human TADG-15 codi
22	430	13.0	3147	22	AAH23609	Human TADG-15 anti
23	430	13.0	2955	21	AAH88492	Human matrixphase (
24	430	13.0	3142	22	AAE28099	Human membrane-ty
25	430	13.0	3149	21	AAH84493	Human matrixphase c
26	430	13.0	3152	22	AAH57431	Human intestine ce
27	430	13.0	3159	21	AAK37657	Human peptidase. H
28	430	12.8	2756	21	AACT7957	Human cancer assoc
29	429	12.7	3115	22	ABK08672	Human membrane-ty
30	429	12.7	3115	22	AAH99565	Human protein enco
31	429	12.5	3448	18	AAV06055	Singapore horseho
32	429	12.5	3448	20	AAK20396	Horsehoes crab fac
33	429	12.5	3448	20	AAK20396	Horsehoes crab fac
34	429	12.5	3448	21	AAK22759	Recombinant 5' tru
35	429	12.5	3448	22	AAK22759	Recombinant 5' tru
36	429	12.5	3448	22	AAK22759	Horsehoes crab fac
37	429	12.5	3448	22	AAK22759	Horsehoes crab fac
38	429	12.5	3448	22	AAK22759	Horsehoes crab fac
39	429	12.5	3448	22	AAK22759	Horsehoes crab fac
40	429	12.5	3448	22	AAK22759	Horsehoes crab fac
41	429	12.5	3448	22	AAK22759	Horsehoes crab fac
42	429	12.5	3448	22	AAK22759	Horsehoes crab fac
43	429	12.5	3448	22	AAK22759	Horsehoes crab fac
44	429	12.5	3448	22	AAK22759	Horsehoes crab fac
45	429	12.5	3448	22	AAK22759	Horsehoes crab fac
46	429	12.5	3448	22	AAK22759	Horsehoes crab fac
47	429	12.5	3448	22	AAK22759	Horsehoes crab fac
48	429	12.5	3448	22	AAK22759	Horsehoes crab fac
49	429	12.5	3448	22	AAK22759	Horsehoes crab fac
50	429	12.5	3448	22	AAK22759	Horsehoes crab fac
51	429	12.5	3448	22	AAK22759	Horsehoes crab fac
52	429	12.5	3448	22	AAK22759	Horsehoes crab fac
53	429	12.5	3448	22	AAK22759	Horsehoes crab fac
54	429	12.5	3448	22	AAK22759	Horsehoes crab fac
55	429	12.5	3448	22	AAK22759	Horsehoes crab fac
56	429	12.5	3448	22	AAK22759	Horsehoes crab fac
57	429	12.5	3448	22	AAK22759	Horsehoes crab fac
58	429	12.5	3448	22	AAK22759	Horsehoes crab fac
59	429	12.5	3448	22	AAK22759	Horsehoes crab fac
60	429	12.5	3448	22	AAK22759	Horsehoes crab fac
61	429	12.5	3448	22	AAK22759	Horsehoes crab fac
62	429	12.5	3448	22	AAK22759	Horsehoes crab fac
63	429	12.5	3448	22	AAK22759	Horsehoes crab fac
64	429	12.5	3448	22	AAK22759	Horsehoes crab fac
65	429	12.5	3448	22	AAK22759	Horsehoes crab fac
66	429	12.5	3448	22	AAK22759	Horsehoes crab fac
67	429	12.5	3448	22	AAK22759	Horsehoes crab fac

```
68 411 10.9 3919 19 AAV10355 Human tollold-like
69 411 10.9 5145 20 AAV45185 Nucleotide sequenc
70 410 10.9 2792 19 AAX28626 Nucleotide sequenc
71 410 10.9 13928 22 AAH77500 Haemophililla B gene
72 408.5 10.9 2381 15 AAQ70104 Bovine enterokinas
73 408 10.9 1467 19 AAV56776 Human Factor X gen
74 408 10.9 1467 19 AAV56821 Human Factor X gen
75 408 10.9 1467 22 AAF59409 Human factor X nuc

ALIGNMENTS

RESULT 1
AAD24224
AAD24224 standard; cDNA: 2472 BP.
AC AAD24224;
DT 17-MAY-2002 (first entry)

Human MASP-2 cDNA clone phl-4.
Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
complement fixation; infection; microbe; retrovirus; HIV; abortion;
human immunodeficiency virus; pathogenic bacteria;
inflammatory disorder; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX 5'UTR 1..36
XX CDS 37..2094
XX FT /tag= a
XX FT /product= "MASP-2 precursor protein"
XX FT /transl_except= (pos:499..500, aa:His)
XX FT /transl_except= (pos:501, aa:Cys)
XX FT /transl_except= (pos:925..927, aa:His)
XX FT /note= "The last two translation exceptions occur
XX only while decoding the alternative version of the
XX MASP-2 protein (AAE14568)"
XX FT sig_peptide 37..81
XX FT /tag= c
XX FT mat_peptide 82..2091
XX FT /tag= d
XX FT /product= "Mature MASP-2 protein"
XX FT 2095..2472
XX 3'UTR /tag= e
XX 0200206460-A2.
XX 24-JAN-2002.
XX 13-JUL-2001; 2001WO-DK00499.
XX 13-JUL-2000; 2000DK-0001089.
XX 01-JUN-2001; 2001DK-0000870.
XX (JENS/) JENSENIUS J C.
XX (THIE/) THIEL S.
XX jensenius JC, Thiel S;
XX WPI: 2002-179791/23.
XX P-PSDB: AAE14564, AAE14568.
XX use of a polypeptide comprising amino acid sequence derived from
XX mannan-binding lectin associated serine protease-2 (MASP-2) for
XX producing pharmaceutical composition, to treat bacterial, fungal, viral
XX infections
XX Example 1; Fig 6; 76pp; English.
XX
```

```
XX The invention relates to use of a polypeptide derived from
XX mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
XX producing a pharmaceutical composition. MASP-2 is a complement-
XX fixing enzyme and involved in lectin pathway of complement activation.
XX The pharmaceutical composition comprising MASP-2 is useful for
XX treating infections caused by microbes such as fungus, yeast,
XX retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
XX bacteria which are resistant to at least one antibiotic medicament or
XX multiresistant. The polynucleotide encoding MASP-2 is useful for
XX treating patients deficient in MASP-2. The invention also discloses
XX MASP-2 assays which are useful for determination of MASP-2 activity or
XX levels in patients suffering from e.g. infections, inflammatory disorders
XX and spontaneous recurrent abortion. The pharmaceutical composition
XX comprising MASP-2 inhibitor is useful for treating inflammatory
XX disorders. The present sequence is human MASP-2 cDNA.
XX
XX SQ Sequence 2472 BP; 627 A; 645 C; 617 G; 583 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3,559-296 Length: 2472
XX Score: 3738.50 Matches: 685
XX Percent Similarity: 99.85 Conservative: 0
XX Best Local Similarity: 99.85 Mismatches: 0
XX Query Match: 99.48% Indels: 1
XX DB: 24 Gaps: 1
XX
XX US-09-874-198-2 (1-686) x AAD24224 (1-2472)
XX
XX 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
XX 37 ATGAGGCTGCTGACCTCCCTGGGGCTTGCTGTGGCTGGTGCCACCCCTTAGCCCG 96
XX
XX 21 LysTrpProGluProValAPheGlyArgLeuAlaSerProGlyPheProGlyTyrAla 40
XX 97 AAGTGGCTGAAACCTGTGGGGCCCTGGCATCCCGGCTTTCAGGGGAGTATGCC 156
XX
XX 41 AsnAspGlnGluArgTrpTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
XX 157 AATGACACGAGGCGGGCTGGACCTGACTGCACCCCGGCTACCGCTCCCTTAC 216
XX
XX 61 PheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAspPheValylLeuSer 80
XX 217 TTCACCCACTTCGACCTGGAGCTCTCCACCTCGGAGAGACATTCGCAAGCTGACG 276
XX
XX 81 SerGlyAlaLysValLeuAlaThrLeuGlyGlyGlnGluSerThrAspThrGluAla 100
XX 277 TCGGGGGCCAAAGTGTGGCCACGCTGTGGGAGAGACACACAGCGAGCGGCC 336
XX
XX 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspThrPheArgSerAsp 120
XX 337 CCTGGCAGGACACTTTCCTGCTGGGCTCCAGCTGACATTCCTCCGCTCCGAC 396
XX
XX 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaAlaGluAsp 140
XX 397 TACTCCACGAGAAGCGTTACAGGGGTTGAGAGCCTTGATGACCGAGCATTTGAC 456
XX
XX 141 GluGlyGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHisLeu 160
XX 457 GAGTGCAGGTGGCCCGGAGAGGGCCCACTCGGACCAACAC---CACAAACCACTG 513
XX
XX 161 GlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgTyrCys 180
XX 514 GCGGTTTCTACTGCTCTCCGCCGAGGCTACGTTCTGACCCGTAACAAGCCACCTGC 573
XX
XX 181 SerAlaLeuGlySerGlyGlnValPheThrGlnArgSerGlyLeuLeuSerSerProGly 200
XX 574 TCAGCCCTGTCGTCGCGGACAGGCTTCACCCAGAGGTTGGGAGGCTCAGACGCCCTGA 633
XX
XX 201 TyrProArgProTyrProLysLeuSerSerCysThrTyrSerTyrLeuGluGluGly 220
XX 634 TACCCAGCGCGGTATCCAAACTCCAGTTGACTTACAGCATCAGCCTGGAGGAGGG 693
XX
```

QY 221 PheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeu 240
 Db 694 TTCAGTGTATCTGGACCTTGTGGAGTCTTGATGTGGAGACACACCTGTAACCTG 753
 QY 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGly 260
 Db 754 TGCCCTACGACTTCTCAAGATTCAAAACAGACAGAGAAAGAACATGCGCCATTCTGTGGG 813
 QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr 280
 Db 814 AAGACATTCGCCACAGATTCGAACAAAGACACAGCGGTGACCATCCTTTGTGACA 873
 QY 281 AspGluSerGlyAspHisThrGlyTrpLysIleHisTyrThrSerThrAlaGlnProCys 300
 Db 874 GATGAATCAGAGACACACAGCGCTGGAAGATCCACTACACGACACAGCGGCGCTTGC 933
 QY 301 ProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIleLeu 320
 Db 934 CCTTATCCGATGGCGCCACTAATGGCCACGTTTCACTGTGCACGCAAAATACATCCTG 993
 QY 321 LysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlyHisLeuPro 340
 Db 994 AAGACACCTTCTCCACTTTTTCGAGACTGGCTATGAGCTTCTGCAAGGTCACTTGCC 1053
 QY 341 LeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspArgPheMetProAla 360
 Db 1054 CTGAATTCCTTACTGACGTTGTGCAAAAGATGATCTTGGAGCGGCGCAATGCCCGCG 1113
 QY 361 CysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluTyrIle 380
 Db 1114 TGCAGCATTTGTGACTGTGGCCCTCTGTGATGATCACTCCGAGCGGAGAGTACATC 1173
 QY 381 ThrGlyProGlyValThrTrpLysAlaValIleGluThrSerCysGluGluThrPhe 400
 Db 1174 ACAGGCTCCTGGAGTACACCACTCAAAAGCTGTATTCAGTACAGCTGTGAGAGACCTTC 1233
 QY 401 TyrThrMetLysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheTrpThrSer 420
 Db 1234 TACACATGTAAGTGAATGATGTGTAATATGTGTGAGCGTGAATGATTCGACGAGC 1293
 QY 421 SerLysGlyGluLysSerLeuProValCysGluProValCysGlyLeuSerAlaArgThr 440
 Db 1294 TCCAAAGGAGAAACATCCTCCAGCTGTGACCTGTTGTGACTATCAGCCGCCACA 1353
 QY 441 ThrGlyArgIleTyrGlyGlyGlnLysAlaLysProGlyAspPheProTrpGlnVal 460
 Db 354 ACAGGAGGCGTATATATGAGGAGGCAAAAGGCAAAACCTGTGTATTTCTTGTGCAAGTC 1413
 QY 461 LeuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeuLeuTyrAspAsnTrpValLeuThr 480
 Db 1414 CTGATATTAGGTGGAAACACAGCAGCGAGTGCCTTTATATGACAACTGGTCCCTAACA 1473
 QY 481 AlaAlaHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAspIleArgMetGly 500
 Db 1474 GCGGCTCAGCCCTCATGTGCAAAACATGATGATCCGCCCTGCAATTCGAATGGGC 1533
 QY 501 ThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrpSerGlnAlaValPheIleHis 520
 Db 1534 ACCCTAAAGAGTATACCTCATATACACAGCCTGGCTGTAAGCTGTTTATATACAT 1593
 QY 521 GluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsn 540
 Db 1594 GAAGGTTATACATCATGTGCTTGGACAAATGACATACACTGATTAATGATGAATAC 1653
 QY 541 LysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSer 560
 Db 1654 AAGGTGTATCAATATGCAACATCAGCCTATTGTCTGCCAAGAAAGAACGTCGATCC 1713
 QY 561 PheMetArgTrpAspAspIleGlyThrAlaSerGlyTrpGlyLeuThrGlnArgGlyPhe 580
 Db 1714 TTTATATGAGACAGATACATTCGCACTGCTGAGTGGGATTAACCCCAAGGGGTTT 1773
 QY 581 LeuAlaArgAsnLeuMetTyrValAspIleProIleValAspHisGlnLysCysThrAla 600

Db 1774 CTTGCTAGGAATCATATGATATGTCGACATACGATTTGTTGACCATCAAAATGTACTGCT 1833
 QY 601 AlaTyrGluLysProProTyrProArgGlySerValThrAlaAsnMetLeuCysAlaGly 620
 Db 1834 GCATATGAAAGGACCCCATCTCCAGGGGGAAGTGTAACTGTCAATGCTTTTGTGCTGGC 1893
 QY 621 LeuGluSerGlyLysAspSerCysArgGlyAspSerGlyLysAlaLeuValPheLeu 640
 Db 1894 TTAGAAAGTGGGGGCAAGACACACTGACAGGTGACACAGCGGAGCGCACTGTGTTCTTA 1953
 QY 641 AspSerGluThrGluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsnCys 660
 Db 1954 GATATGTAACACAGAGAGGTGTTGTGGAGAAATGTCCTGGGGTTCATGAATTTGT 2013
 QY 661 GlyGluAlaGlyLysIleTyrGlyValTyrThrLysValIleAsnTyrIleProTrpIleGlu 680
 Db 2014 GGGGAAGCAGCTCAGTATGAGTCTACACAAAGTTATTAATATTCCTCTGATCGAG 2073
 QY 681 AsnIleIleSerAspPhe 686
 Db 2074 AACATTAATTAATGATTTT 2091
 RESULT 2
 AAS85321
 ID AAS85321 standard; cDNA: 3106 BP.
 XX
 AC AAS85321:
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #21125.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN MO200175067-A2. *last date*
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217 *last date*
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSBQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG21134.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX Claim 1: SEQ ID NO 21125; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

Sequence 3106 BP; 848 A; 846 C; 798 G; 614 T; 0 other;

874-198-2 (1-686) x AAS85321 (1-3106)

QY 23 ProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAsp 42

QY 43 GINGIARGARGTRPThrLeuThraIaProProGlyTYrArgLeuArgLeuTYrPheThr 62

QY 63 HISPHEASPLEUGLULEUSERHISLEUCYSGLUTYRASPPHEVALLYSLEUSERCGLY 82

83 AlaySValLeuAlaThrLeucGlyGlnGluSerThrAspThrGluArgAlaProGly 102

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

db 2636 CAGGTGCCCCGGAGAGGGCCCCACCTGCGACCACTGCCACAACCACTGGCGGT 2695

Db 2696 TTCTACTGCTCCTGCGCGGAGGCTACGTCCTGCACCGTAACAAGCGCACCTGCTCAGCC 2755

DB 2/56 CIGIGCCTCCGGCCAGGTCCTCACCAGAGGCTCTGGGGAGCTCAGCAGGCCCTGAATACCA 2815

[illegible]

RESULT 3

AC AAH43512;

DT 13-DEC-2001 (first entry)

CDNA encoding protease PRTS-1.

KW Human; protease; PRTS; gastrointestinal disorder; peptic esophagitis;

KW hypertensive heart disease; myocardial infarction; autoimmune disorder

KW arteriosclerosis; cancer; epithelial disorder; eczema; endometriosis;

XX

Homo sapiens.

FH	key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT / *tag = a

XX

XX

XX

XX XX

PR 30-MAR-2000; 2000US-193182P

PR 20-APR-2000; 2000US-199022P

XX

XX

PI Au-Young J, Bandman O, La

PI Hafalia A, Lu Y, Patterson

WPI; 2001-611509/70.

XX
XX

PT autoimmune disorders, comprising

XX
XX

2
. |

n

j

.
o
o
o
o
.
o
o
|

XX
XX
XX

to -11. These proteases can

expression of functional pro-

gastrointestinal, cardiovascular

proliferative (arteriosclerosis)

disorders. The protein encoded

XX

Db 2057 GACACGTGCTGGAGATACGGGTGGGGCCTTGTCTTGTATGACTTGACCCAGCGC 2116
 QY 647 TrpPheValGlyIleValSerTrpGlySer---MetSnySGlyIuAlaGlyGln 665
 Db 2117 TGGGTGGTGCAGAGCGCTGGCTCGGGGACCTGAGAAATGCGGACCAAGCAGGTG 2176
 QY 666 TyrGlyValTyrThrIysValIleAsnTyrIleProTrpIle 679
 Db 2177 TATGGAGCTACACAAAGCTGTCCAAATTAATGCTGGCTGGGTG 2218
 RESULT 4
 ID AAF83985
 AC AAF83985 standard; DNA; 3895 BP.
 XX AAF83985;
 AC
 XX 06-AUG-2001 (first entry)
 DT
 XX Human serine protease MASP-3 encoding DNA.
 DE
 XX Human-binding lectin associated serine protease-3; MASP-3; MBL; human;
 KW immunomodulator; neuroprotective; anticonvulsive; antithrombotic;
 KW antidiabetic; antiatherosclerotic; ds.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 91..227
 FT /*tag= a
 FT /product= "MASP-3"
 XX
 PN MO200140451-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 30-NOV-2000; 2000MO-DK00659.
 XX
 PR 02-DEC-1999; 99DK-0001721.
 PR 21-JUL-2000; 2000DK-0001126.
 XX
 PA (JENS/) JENSENIUS J C.
 PA (THIE/) THIEL S.
 XX
 PI Jensenius JC, Thiel S;
 XX MPI, 2001-374820/39.
 DE
 XX P-PSDB; AAB85060.
 XX
 XX novel pure mannan-binding lectin associated serine protease
 XX polypeptides and polynucleotides encoding the protein, used for
 PT treating reoxygenated ischemic tissues, mannan-binding lectin
 PT deficiency, multiple sclerosis
 XX
 PS Claim 18; Page 88-96; 99pp; English.
 XX
 CC The invention relates to a mannan-binding lectin (MBL) associated serine
 CC protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on
 CC the complement activation particularly when bound to MBL/MASP-2 complexes
 CC and directly activates complement system through binding to MBL. The
 CC MASP-3 polypeptides, polynucleotides and modulators are useful for
 CC preparing a pharmaceutical composition for treating aberrant MASP-3
 CC activity such as infections, cancer, MBL-deficiency, disorders of the
 CC immune system and reproductive system. The MASP-3 polypeptide is also
 CC used for treating diseases associated with human immunodeficiency virus,
 CC multiple sclerosis, myasthenia gravis, epilepsy, rheumatoid arthritis,
 CC vasculitis, autoimmune hemolytic anemia, Crohn's disease, asthma,
 CC diabetes, psoriasis, multiple myeloma, atherosclerosis etc. It is also
 CC useful for inhibiting activation of C4 complement by inhibiting the MBL
 CC pathway, inhibiting MASP-2 activity, inhibiting or treating an
 CC inflammatory condition related to complement activation through MBL/MASP
 CC complexes. MASP-3 is also useful for treating an inflammatory condition

CC resulting from an autoimmune condition after acute myocardial infarction
 CC or brain ischemia. It is also useful for treating an individual suffering
 CC from a disorder resulting from an imbalanced cytokine network. The
 CC present sequence represents a DNA encoding a human MASP-3 polypeptide.
 XX

Sequence 3895 BP; 952 A; 1068 C; 987 G; 888 T; 0 other;

Alignment Scores:

pred. No.:	6 44e-121	Length:	3895
Score:	1600.00	Matches:	305
Percent Similarity:	60.228	Conservative:	125
Best Local Similarity:	42.724	Mismatches:	246
Query Match:	42.588	Indels:	38
DB:	22	Gaps:	11

US-09-874-198-2 (1-686) x AAF83985 (1-3895)

QY 1 MetArgLeuLeuThrLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
 Db 91 ATGAGGTGGCTGCTCTCTATATGCTCTGTGCTCTGCTCTGCAAGGCTTCAGCCAC 150
 QY 21 LysTrpPro---GluProValPheGlyArgLeuAlaSerProGlyPheProGlyIuTyr 39
 Db 151 ACCGTGAGCTAAACAATATATGTTGGCCAGATCCAGTCCGCTGTTATCCAGACTCTAT 210
 QY 40 AlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeu 59
 Db 211 CCCAGTATTCAGAGGTGACTTGGAAATTCACCTGCTCCAGATGGTTTGGATCAAGCTT 270
 QY 60 TyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAspPheValLysLeu 79
 Db 271 TACTTCATGCACTTCAATCAATCGAATCCCTCTACTTGTGGAATATGATGATGGAAGTGA 330
 QY 80 SerSerGlyAlaLysValLeuAlaThrLeuGlyGlnGluSerHisThrAspThrGluArg 99
 Db 331 GAAACTGAGGACCAAGGTGCGGCAACCTTGTGGAGGAGACCAACACAGACAGACAG 390
 QY 100 AlaProGlyLysAspThrPheThrLeuGlySerSerLeuGlySerSerLeuThrPheArgSer 119
 Db 391 ACTCCCGGCCAGAGAGGTGGTCTCTCCCTCGCTCTTCTATGTCATCTTCCTCCGCTCA 450
 QY 120 AspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaLaGluAspIle 139
 Db 451 GATTTCCTCCAAATGAGAGCGTTTCACAGGCTTGTAGCCACATGCTGCTGGATGATG 510
 QY 140 AspGluGlyGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHis 159
 Db 511 GACGAGTGCAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 570
 QY 160 LeuGlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgTyr 179
 Db 571 ATTGGGGGCTACACTGCTCTCTCCCTCGCTTCTGCTATCTCTCCACACAGACAGAGGAG 630
 QY 180 CysSerAlaLeuGlySerSerGlyGlnValPheThrGlnArgSerGlyLysLeuSerSerPro 199
 Db 631 TGCCGAGTGGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 690
 QY 200 GluTyrProArgProTyrProLysLeuSerSerCysThrTyrSerTrpLeuSerLeuGlu 219
 Db 691 GACTTCCCAAAACCTTACCCCAAGAGCTCTGATGCTGCTATACCTCAGACGCGAGGAG 750
 QY 220 GlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGlyThr 239
 Db 751 GGTTTATGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 810
 QY 240 LeuGlyProTyrAspPheLeuLysIleGlnThrAspArgGluGluHisGlyPropheCys 259
 Db 811 CCTGGCCCTATACATCAATCAAGATCAAGTTGGTGGCAAAAGTTTGGGCGCTTCTGT 870
 QY 260 GlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheVal 279
 Db 871 GGAGAGAAAGCCCAAGACCAACCATCACACCCAGAGCCACAGTGTCTGATCTGTTCCAT 930

QY	280	ThrspluSerGlyaspHisphStrHlyrPlyvSIleHStrYthSeThrAlaIlepro	299
Db	931	AGTACAACTCGCAGAGAACCGGGCTGAGGCTCTCATACAGGGCTGCAGAAATGAC	990
QY	300	CysProTyrProMetAlaProPheasnGlyHisValSerProValGlnAlaLysTyrIle	319
Db	991	TGCCAGAGCTACAGGCTCTGTCATGGGAAATGAGAGCCCTCCCAAGCAAGTATTC	1050
QY	320	LeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlyHisLeu	339
Db	1051	TTTCAAACCAACCAAGTCGTCGACACTGTGCACAGGCTCAACAGGCTCAAAAGTCTCAAGAGTAAATGTG	1110
QY	340	ProLeuLysSerPheThrAlaValGlySglnLysAspGlySerThrAspArgProMetPro	359
Db	1111	GAGATGGACACATTCACAGATTGAGTGTCTGAAAGAGTGGAGCTGGAGTACACAGATTTCC	1170
QY	360	AlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluTyr	379
Db	171	ACCTGATAAATTGTAGACTGTACAGCCCGCAGAGAGCTGGAAACAGGGCTGATCACTTC	1230
QY	380	IleHcGlyProGluValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThr	399
Db	1231	TCATCAAGGAACACCTCCACACATCAAGTGTGAGATCAATTAATCTCTGTCAGAGGCC	1290
QY	400	PheTyrThrMet---LysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheTrp	418
Db	1291	TATTACAGATGCTCAACAATTAACACAGGTATATATACCTGTTCGCCCAAGGACTGTGG	1350
QY	419	ThrSerSerLysGlyLysSerLeuProValCysGluProValCysGlyLeuSerAla	438
Db	1351	ATGAAATTAATATTGGGGAGAACCTTACCCACTGCTCTCCAGATGTGTGTACGGCTCC	1410
QY	439	ArgThrThrGlyGly-----ArgIleTyrGlyGlyGlnLysAlaLysProGlyLysP	455
Db	1411	CGCTCCCTGCCCAAGCTGTGTCAAGAGATCAATTGGGGCCGAAATGTGAGCCTGCGCTC	1470
QY	456	PheProTrpGlnAlleuIleLeuGlyLysThrAla-----	468
Db	1471	TTCCCGTGGCAGGCCCTGATAGTGGTGGAGGACCTCGAGAGTCCCAATATGACAAATGG	1530
QY	469	-----AlaGlyAlaLeuLeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyr	486
Db	1531	TTTGGGAGTGGGGCCCTGCTCTGTGGTCTGTGATCTTACAGCAGCTCATGTGCTGCC	1590
QY	487	GluGlnLysHisAspAlaSerAlaLeuAspIle--ArgMetGlyThrLeuLysArgLeu	505
Db	591	TCCACGGGTAGAGACACACAGGTGATACCAAGTCTCCAAAGACATGTACCGTACTGTG	1650
QY	506	SerProHisTyrThrGlnAlaTrpSerGluAla-----ValPhe	518
Db	1651	GGCTTGCTATGTGGGAGACAAATATGGGGGAGTCAACAGCTCAGTCCGCGAGTGGTG	1710
QY	519	IleHisGluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeu	538
Db	1711	CTCCACCAACACTTC---AACATCCAAACTACAAACAGATATAGCTGTGGTGAAGTGTG	1767
QY	539	AsnAsnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAla	558
Db	1768	CAGAGCGCTGTGCCCTGGGAGCCACACGTTATGCGTGCCTGCGTCCCAAGCGTTAGAGCT	1827
QY	559	GluSerPheMetArgThrAspAspIleGlyTyrAlaSerGlyTyrPglyLeuThrGlnArg	578
Db	1828	GAAAGC---CGGGCCCCCACCACATGCTGGCGCTGTGGCGCGGTGGGAGTCTCCAAATCC	1884
QY	579	GlyPhe-----LeuAlaArgAsnLeuMet	586
Db	1885	AATGTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1944
QY	587	TyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProPro	606
Db	1945	TATGTCAAGTATCCCGGCGGCTCACCGCTGAGTGGAATAAATGATATGATGATGATGATGATG	2004
QY	607	TyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGlyLys	626

Db	2005	---	GGCAATTACAGCGCTCACGAGAAACATGTTCTGTGCTGGCTACTACGAGGCGCGCAA	2061
Qy	627	AspSerCysArgIlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGluArg	646	
Db	2062	GACACGCGCTTGAGAGTATAGCGGTGGCGGCTTTGTGCATCTTGATGACGTTGACGACGCGC	2121	
Qy	647	TrpPheValIleGlyIleValSerTrpGlySer---MetAsnCysGlyGluIaGlyCln	665	
Db	2122	TGGGTGTCGACAGCGCTGCTGCTCTCGGGGGGACCTGAGAAATGCGGACGACAGCAGTC	2181	
Qy	666	TyrGlyValIlyThrIlyValIleAsnTyrIleProTrpIle	679	
Db	2182	TATGAGCTCTACAAAGGCTCTCAATTACGTGAGCGGTG	2223	
RESULT 5				
AXX24297				
AC	AXX24297	standard:	cdna: 2492 BP.	
XX	AXX24297:			
XX	07-JUN-1999	(first entry)		
DE	Human complement component 1, subcomponent r (c1r) cdna.			
XX	Genecalling: quantitative expression analysis; oligo-poisoning;			
KW	c1r gene; complement component 1 subcomponent r; human; ss.			
KM				
XX				
OS	Homo sapiens.			
XX				
FT	Key	Location/Qualifiers		
FT	misc_feature	1875..1880		
FT	/*tag= a			
FT	/note= "BspHI recognition sequence"			
FT	misc_feature	2187..2192		
FT	/*tag= b			
FT	/note= "EcoRI recognition sequence"			
XX	WO9907896-A2.			
PD	18-FEB-1999.			
XX				
PF	07-AUG-1998:	98WO-US16548.		
XX				
PR	07-AUG-1997:	97US-0054887.		
XX				
PA	(CURA-) CURAGEN CORP.			
XX	Deem MW, Rothberg JM, Slompson JM;			
PI	WPI; 1999-167456/14.			
DR				
XX				
PT	Detection and confirmation of a nucleotide sequence - by primer			
PT	hybridisation to known terminal sequences and unidentified internal			
PT	sequences, used to indicate, e.g. risk of neuroblastoma			
XX				
PS	Disclosure; Fig 4: 52pp; English.			
XX				
CC	This is the nucleotide sequence of human complement component 1,			
CC	subcomponent r (c1r) cdna generated from c1r poly(A)+ mRNA. It was			
CC	utilised to demonstrate an "oligo-poisoning" method of confirming			
CC	the identity of a nucleic acid. The c1r cdna was digested with			
CC	BspHI and EcoRI to produce a 318 bp fragment. This was isolated			
CC	and subjected to PCR amplification using primers J (see AAX24298) or			
CC	R (see AAX24299), which are complementary to the EcoRI and BspHI			
CC	recognition sites of c1r. The electrophoretic banding patterns and			
CC	mobilities of the amplification products produced by J-primer and			
CC	R-primer PCRs were used to generate electropherograms. Utilising			
CC	the results obtained in this first, genecalling (TM), PCR reaction,			
CC	sequence confirmation was performed by oligo-poisoning. 2			
CC	poisoning primers (see AAX24300 and AAX24301) were generated which had			
CC	sequences corresponding to the last 6 nucleotides of the J and R			
CC	primers, respectively, followed by 16 nucleotides corresponding to			
CC				

CC the c1R cDNA sequence. PCR amplification was performed, and
 CC electropherograms were generated. An approximate 3-fold reduction
 CC in signal corresponding to the 318 bp fragment was observed, thus
 CC confirming the nucleic acid identity derived from the original
 CC Genecalling methodology. The general method of the invention
 CC provides direct confirmation that nucleic acid fragments,
 CC possessing putatively identified sequences which have been predicted
 CC to generate observed Genecalling signals, are actually present
 CC within the sample generating the signal. It provides rapid,
 CC economical, quantitative and highly specific determination or
 CC classification of DNA sequences, especially genomic DNA or cDNA
 CC sequences, in either arrays of single sequence clones or mixtures
 CC of sequences such as can be derived from tissue samples, without
 CC actually sequencing the DNA. The methods can be used to detect the
 CC presence and expression of certain genes or their particular
 CC alleles, which can be indicative of risk factors for diseases, e.g.
 CC Huntington's disease and ataxia telangiectasia, as well as several
 CC cancers, e.g. neuroblastoma. The methods can also be used to
 CC determine the presence and classification of foreign pathogens
 CC which are difficult or impossible to culture in vitro, but express
 CC their own genes. They may also be used to assess disease
 CC progression, and for pharmacogenomic analysis of drug action and
 CC efficiency.

XX Sequence 2492 BP; 619 A; 679 C; 662 G; 532 T; 0 other;

Alignment Scores:

Score:	8,76e-106	Length:	2492
Percent Similarity:	1412.00	Matches:	301
Best Local Similarity:	56.73%	Conservative:	108
Query Match:	41.75%	Mismatches:	248
	37.57%	Indels:	64
DB:	20	Gaps:	22

US-09-874-198-2 (1-686) x AAX24297 (1-2492)

QY 1 MetArgLeuLeuThrLeuLeu-----GlyLeuLeuCys-----GlySerValAla 15
 DB 64 ATGGGCTCTGTGACCTCGTGCCGCGCCCTGTTCTGCAGGCGAGAGGCTCCATT--- 120
 QY 16 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 35
 DB 121 ---CCCATC-----CCTCAGAAAGTATTGTGGGAGTGACTTCCCTCTGTTTC 165
 QY 36 ProGlyLeuTrpAlaAsnAspGlnGluArgTrpThrLeuThrAlaProGlyTyr 55
 DB 166 CCCAAGCCTTACCCCAACACTTGTGAACAACACCTGTGATCAAGTCCCAAGCGGATAC 225
 QY 56 ArgLeuArgLeuTrpThrHisPheAspLeuGluLeuSerHisLeuCysGluTyrAsp 75
 DB 226 AGGGTGAAGCTCGTCTCCAGACAGTTTGACCTTGAGGCTTCTGAAGCTGCTCTATGAT 285
 QY 76 PheValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGluSerThr 95
 DB 286 TATGTCAAGATCTGTGATAGAAAGCCTGGGAGGTTCTGTGGCAACTGGGTTCT 345
 QY 96 AspThrGluArgAlaProGlyLysAspThrPheTrpSerLeuGlySerSerLeuAspIle 115
 DB 346 CCAATGGGCAACCCCGCGGAAGAAAGAAATTATGTGCCAAGGAAACAAAGATCTGCTG 405
 QY 116 ThrPheArgSerAspTrpSerAsnGlnLysPro-----PheThrGlyPhe 130
 DB 406 ACCTTCCACACAGACTCTCCAAAGAGAGATGAGCAATCATGTTCTTCAGAGGCTTC 465
 QY 131 GluAlaPheTrpAlaAlaGluAspIleAspLysCys-----GlnValAlaProGlyGlu 148
 DB 466 CTGGCGCTACTACCAAGGCTGGAGCTTGATGAATGTCTTCCGAGCAACAATCAGGGGAG 525
 QY 149 -----AlaProTrpCysAspHisHisCysHisAsnHisLeuGlyLysPheTyrCys 165
 DB 526 GAGGATCCCAAGCCCAAGTGGCAGACCTGTGTACAACTAGTGGAGGCTACTTGTGT 585
 QY 166 SerCysArgAlaGluTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSer 185

DB 586 TCTGCGCGTCCAGGCTATGAGCTTCAAGAGACAGAGCATCTCCAGCGCTGAGTCAGC 645
 QY 186 GlyLysValPheThrGlnAlaGlySerGlyGluLeuSerSerProGluTrpProArgProTyr 205
 DB 646 AGCGATGTGACACGGGAGGATCAGCTACATCTCCAGCTTGAGATCCCTCGCTTAC 705
 QY 206 ProLysLeuSerSerCysTrpTrpSerIleSerLeuGlnGluLysPheSerValIleLeu 225
 DB 706 CCCCTCACTGCGCTCAGCTCACTACAGATCCGGGTGAGGGGGCCCTACCCCTGACCTC 765
 QY 226 AspPheValLysSerPheAspValGluThrHisProGluThrLeuCysProTyrAspPhe 245
 DB 766 AAGTCTCGAGCCTTGTGATTTGATGAGACACAGCAATACATGCTCCCTATGACACAG 825
 QY 246 LeuLysIleGlnThrAspArgGluGlnHisGlyProPheGlySerGlyLysThrLeuProHis 265
 DB 826 CTACAGATCTATGCCAAGCGGAAACATTTGGCAGTTCTGTGGGAAGCAAGGCCCC 885
 QY 266 ArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGluSerGlyAsp 285
 DB 886 GACCTGCACACACAGACATGCTGTGATCTGCTGTTCTTCACAGATGATGGGGGAC 945
 QY 286 HisThrGlyTrpLysIleHisTyrThrSerThrAlaGlnProCysProTyrProMetAla 305
 DB 946 AGCCGGGCTGGAAGCTGCGCTACACACAGATCATCAAGTCCGCCCAAGACC 1005
 QY 306 ProProAsn---GlyHisValSerProValGlnAlaLysTyrIleLeuLysAspSerPhe 324
 DB 1006 CTAGACAGTTCACCATCATCAACACCAACCTGACCTCAGTACAGTTCCTGACTCTTC 1065
 QY 325 SerIlePheCysGluThrArgIleTyrGluLeuLeuGlnGlyHisLeuProLeuLysSerPhe 344
 DB 1066 ATTGCTACCTGCACAAAGCTTACACAGCTCATAGAGGGAACCAAGGTGCTGATTCTTC 1125
 QY 345 ThrAlaValCysGlnLysAspGlySerTrpAspArgProMetProAlaCysSerIleVal 364
 DB 1126 ACAGCTGTCTGCAGATGATGACAGTGTGACATGTCGCATGCCAGATGCAAGATCAAG 1185
 QY 365 AspCysGlyProProAspAspLeuProSerGlyArgValGluThrIleThrGlyProGly 384
 DB 1186 GACTGTGGGAGCCCGAAACCTGCTTAATGTGACTCCGTTTACACCAACCAATGGGA 1245
 QY 385 ValThrThrTrpTrpAlaValIleGlnTyrSerCysGluGluThrPheTrpThrMetLys 404
 DB 1246 GTGACACCTTACAAAGGCCCTATTCAGTACTGCGCATAGGCCATATTAACAAGATGAC 1305
 QY 405 Val-----AsnAspGlyLysTrpValLysGluAlaAspGlyPhe 417
 DB 1306 ACCAGACCTGGCAGCAGGAGTCTGAGCAAGGGGTGTACCTGACACAGCAGAGGCATTT 1365
 QY 418 Trp---ThrSerSerLysGlyLysSerLeuProValLysGluProValLysGly--- 435
 DB 1366 TGGAAAGATTAACAAGAGGAGAGAG--ATTCTCGGTGCTTGCAGTGTGTGGAG 1422
 QY 436 ---LeuSerAlaArgThrTrpGlyArgIleTyrCysGlyGlnLysAlaLysProGly 454
 DB 1423 CCGGTAAACCCCGTGAACAGAGGCAAGCGCATATATGAGAGGAAAAAGCCAAAGATGGC 1482
 QY 455 AspPheProTrpGlnValLeuIleLeuGlyLysThrAlaAlaGluAlaLeuLeuTyr 474
 DB 1483 AACTTCCCTCGCAGGCTGTCCACCAACATCCAGCGCGCGCGCGCGCTGTGGGC 1542
 QY 475 AspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysHisAspAla----- 492
 DB 1543 GACCGGTGATCTCCACAGCTGACACACCTGTATCCCAAGGAACAAGAGGCAAAAG 1602
 QY 493 ---SerAlaLeuAspIleArgMetGly-----ThrLeuAspArgLeuSer 506
 DB 1603 AAGGCTCTTGTGATGTGCTTGGGCGACACAAATGTGGAAGAGCTCATGAAGCTTAGGA 1662
 QY 507 ProHisTyrThrGlnAlaTrpSerGluAlaValPheIleHisGluLysIleTyrThrHisAsp 526

Db 1663 AATCACCCTCCG-----AGGCTAGCGCTCCACCCGGACTACCTCAGGAT 1710
 QY 527 Ala-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnValIle 544
 Db 1711 GAGTCCACAAATTTGAGGGGAGCATCGCCCTGCTGAGAGCTGGAATATGATCACCCTG 1770
 QY 545 AsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThr 564
 Db 1771 GGTCACAACTCTCCCACTGCTCCCT-----GACAAGCATACCTCTACGACCTG 1824
 QY 565 AspAspIleGlyThrAlaSerIleThrGlyLeuThrGlnArgGlyPheLeuAlaArgAsn 584
 Db 1825 GCGTTGATGGGCTATGTCACATGCTTCGGGCTATGGAGAGAG--ATTGCTCATGAC 1881
 QY 585 LeuMetArgValAspIleProIleValAlaPheHisGlnLysCysThrAlaIleArgLys 604
 Db 1882 CTCAGGTTTCCGCTCGCCCGCTAGCTAATCCACAGCCCTG----- 1923
 QY 605 ProProThrProArgLys-----ValThrAlaAsnMetLeuCysAla 619
 Db 1924 GAGAACTGCTCCGGGAGAAAGATAGATGATGTTCTCTCAAAACATGTTCTGTGCT 1983
 QY 620 GlyLeuGluSerGlyLysAspSerCysArgGlyAspSerGlyAlaLeuValPhe 639
 Db 1984 GGACACCCCATCTTAAAGCAGAGCGCTGCCAGGGGATGAGGGGCTTTTTCAGTA 2043
 QY 640 LeuAspSerIleThrGluArgTrpPheValGlyIleValSerTrpGlySerMetAsn 659
 Db 2044 AGGAGACCCGAACTATGCTGCTGGTGGCCAGCGGCTGCTGCTGGGCT---ATCGG 2100
 QY 660 CysGlyGluAlaGlyLysIleValThrLysValIleAsnTrpIleProTrpIle 679
 Db 2101 TGCAGCAGGGGC-----TATGCTTCTACACCAAGTCTCAACATCAGTGGACTGATC 2154
 QY 680 GLu 680
 Db 2155 AAG 2157
 Db
 RESULT 6
 ID ABA83114
 ID ABA83114 standard; DNM; 2493 BP.
 AC ABA83114;
 XX
 DT 08-FEB-2002 (first entry)
 XX
 Complement component 1 r ovarian tumour marker gene, SEQ ID NO:65.
 Ovarian tumour marker gene; human; overexpression; upregulation;
 epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 identification; serous cystadenoma; borderline serous tumour;
 serous cystadenocarcinoma; mucinous cystadenocarcinoma;
 mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
 undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
 immune response pathway; cell proliferation regulation; protein folding;
 membrane localised; secreted; therapeutic target; cytostatic;
 gene therapy; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200175177-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 03-APR-2001: 2001WO-US10947.
 XX
 PR 03-APR-2000: 2000US-194336P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 XX

DR WPI: 2001-626450/72.
 DR P-PSDB: ABB50288.
 XX
 PT detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian
 PT cancer treatment, by measuring expression level of ovarian tumor marker
 PT gene
 XX
 PS Claim 23: Page 107; 140pp; English.
 XX
 CC The invention relates to methods for diagnosing and prognosing ovarian
 CC tumours in an individual via the detection and measurement of the
 CC expression of an ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
 CC ABA83181 and ABA83183). The methods of the invention are useful for
 CC detecting an ovarian tumour in a patient, for identifying an individual
 CC at increased risk for developing ovarian cancer, in prognostic tests for
 CC assessing the relative severity of ovarian cancer, in tests for
 CC monitoring a patient in remission from ovarian cancer and in tests for
 CC monitoring disease status in a patient being treated for ovarian cancer.
 CC The methods can additionally be used to identify a particular tumour as
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
 CC tumour. The ovarian tumour marker genes of the invention were identified
 CC using SAGE (serial analysis of gene expression) and were found to be
 CC overexpressed in a broad variety of ovarian epithelial tumour cells
 CC relative to normal ovarian epithelial cells. The marker genes are
 CC implicated in immune response pathways, in the regulation of cell
 CC proliferation and in protein folding, and many of these are membrane-
 CC localised or secreted. In addition to their use as diagnostic and
 CC prognostic markers, the ovarian tumour marker genes or their encoded
 CC proteins may be used as therapeutic targets for the treatment and
 CC prevention of ovarian cancer. Sequences ABA83081-ABA83122, ABA83180,
 CC ABA83182 and ABA83184 represent the ovarian tumour marker genes of
 CC the invention.
 CC
 XX
 SO Sequence 2493 BP; 619 A; 680 C; 662 G; 532 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 8, 77e-106 Length: 2493
 Score: 1412.00 Matches: 301
 Percent Similarity: 56.73% Conservative: 108
 Best Local Similarity: 41.75% Mismatches: 248
 Query Match: 37.57% Indels: 64
 DB: Gaps: 22
 US-09-874-198-2 (1-686) x ABA83114 (1-2493)
 QY 1 MetArgLeuLeuThrLeuLeu-----GlyLeuLeuCys-----GlySerValAla 15
 Db 64 ATGTGGCTCTTGAACCCCTCGGCGCGCCCTGTTCGCGAGGCGAGGACCTCCATT-- 120
 QY 16 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 35
 Db 121 ---CCCATC-----CCTCAGAACTTATTGGGAGGAGTGACTTCCCTCTGTTTC 165
 QY 36 ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTrp 55
 Db 166 CCCAAGCCTTACCCCAACACACTTGAACAACACACTGTATCAAGTCCCGGAGTTAC 225
 QY 56 ArgLeuArgLeuTrpPheThrHisPheAspLeuGluLeuSerHisLeuCysGluTrpAsp 75
 Db 226 AGGCTGAAGCTCGTTCGACGAGCTTTGACGAGCTTCTGAAGGCTCTCTATGAT 285
 QY 76 PheValLysLeuSerSerGlyAlaLysValIleAlaThrLeuCysGlyGlnGluSerThr 95
 Db 286 TATGTCAAGATCTCTGTGATTAAGAAAGCCTGGGAGAGTTCGTGGGCAACTGGCTTC 345
 QY 96 AspThrGluArgAlaProGlyLysAspThrPheTrpSerLeuGluLysSerLeuAspIle 115

Db 346 CCACTGGGCAACCCCGGGGAAGAAAGAAATTATGTCCCAAGGGAACAGATGCTGCTG 405
 QY 116 ThrPheArgSerAspTyrSerAsnGluLysPro-----PheThrGlyPhe 130
 Db 406 ACCTTCACACAGACTTCTCCACGAGGAGAAATGGACCATCATGTTCTCTCAAGGGCTTC 465
 QY 131 GluAlaPheTyrAlaAlaGluAspIleAspIleCys-----GluAlaAlaProGlyGlu 148
 Db 466 CTGGCTACTACCAAGCTGTGGACCTTGATGATGTGCTCCCGGAGCAATATCGGGAG 525
 QY 149 -----AlaProThrCysAspHisHisCysHisAsnHisLeuGlyGlyPheTyrCys 165
 Db 526 GAGAGTCCCGACCGCCGCTCCACACCTGTGTCAACAACATGCTGGAGCTTACTTGTGT 585
 QY 166 SerCysArgAlaGlyTyrValIleuHisArgAsnLysArgThrCysSerAlaLeuCysSer 185
 Db 586 TCTCGCCGCTCAGGCTTAGAGCTTCAGGAAGACAGGCAATTCCTCTCCAGGCTGAGTGGACC 645
 QY 186 GlyAlaValPheThrGlnArgSerGlyLysLeuSerSerProGluTyrProArgProTyr 205
 Db 646 AGCAGAGCTAGACAGGAGGATCGAGTACATCTCCAGCCGAGATACCCCTGGCTCTAC 705
 QY 206 ProLysLeuSerCysThrTyrSerIleSerLeuGluGluGlyPheSerValIleLeu 225
 Db 706 CCCCCTACCTGCGCTCACTACAGCATCGGGTGGAGGGGCGCTGCACCTGCACCTC 765
 QY 226 AspPheValGluSerPheAspValGluThrHisProGluThrLeuGlyCysProTyrAspPhe 245
 Db 766 AAGTCTCGACCTTTTGATTTGATGATGACCCAGCAAGTACACTGCCCTATAGACAG 825
 QY 246 LeuLysIleGlnThrAspArgGluGluHisGlyProPheCysGlyLysThrLeuProHis 265
 Db 826 CTACAGATCTATGCGACGGAAGAACATGGGCGAGCTTGTGGAGAAAGCAAGGCCCCC 885
 QY 266 ArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGluSerGlyAsp 285
 Db 886 GACCTCACACCAACAGCATGCTGTGATCTGCTTTCACAGATGAGTGGGGGAC 945
 QY 286 HisThrGlyTyrLysIleHisTyrThrSerThrAlaGlnProCysProTyrProMetAla 305
 Db 946 AGCGGGGGCTGGAGGCTGCGCTACACCAAGCATCATCAAGTCCCCCAAGCCCAAGACC 1005
 QY 306 ProProAsn---GlyHisValSerProValGlnAlaLysTyrIleLeuLysAspSerPhe 324
 Db 1006 CTAGACAGATTACCATCATCCAGAACCTGACGCTCAGTACAGCATGCCGACTACTTC 1065
 QY 325 SerIlePheCysGluThrGlyTyrGluLeuGluGlnGlyHisLeuProLeuLysSerPhe 344
 Db 066 ATTGCTACCTGCAGCAAGGCTTACAGCTCATATAGAGGGAACAGAGTGCATTCCTTC 1125
 QY 345 ThrAlaValCysGlnLysAspGlySerThrAspArgProMetProAlaCysSerIleVal 364
 Db 1126 ACAGCTGTCTGCGAGATGATGGACATGCGATGCTCCATGCCCAAGATGCAAGATGAAG 1185
 QY 365 AspCysGlyProProAspAspLeuProSerGlyArgValGluTyrIleThrGlyProGly 384
 Db 1186 GACTGTGGGAGCGCCGAACCTGCTTAATGTGACTTCGCTTACACCAACCAAGATGGGA 1245
 QY 385 ValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMetLys 404
 Db 1246 GTGAACACCTACAGAGGCGTATCCACTACTAGCCATAGCCATATTACAGATTCAG 1305
 QY 405 Val-----AsnAspGlyLysTyrValCysGluAlaAspGlyPhe 417
 Db 1306 ACCAGAGCTGGACAGGAGGTCTGACCAAGGGGTGTACACCTGCACAGCAAGGGCATTT 1365
 QY 418 Trp---ThrSerSerLysGlyLysLysSerLeuProValCysGluProValCysGly--- 435
 Db 1366 TGGAGAATGAACAGAGGAGGAGAAAG---ATTCTCGGTGGCTTGCACAGTGTGGGAAG 1422
 QY 436 ---LeuSerAlaArgThrThrGlyGlyArgIleTyrGlyGlyGlnLysAlaLysProGly 454
 Db 1423 CCGGTGAACCCCGTGAACAGAGGACGCAATATCGAGGGCAAAAAGCCAAAGATGGGC 1482

QY 455 AspPheProTrpGlnValLeuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeuTyr 474
 Db 1483 AACTTCCCTCGGACAGGTGTTCCACCAATCCAGCGGGCGGGGGCCCTGCTGGGGC 1542
 QY 475 AspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysHisAspAla----- 492
 Db 1543 GACCGCTGATCTCTACAGCTGCCACACCTGTATCCCAAGGAACAGCAAGCGCAAGC 1602
 QY 493 ---SerAlaLeuAspIleArgMetGly-----ThrLeuLysArgLeuSer 506
 Db 1603 AACGCTCTTGGATGTGTTCTGGGCCACACAATGTGAAGAGCTCATGAAGCTAGGA 1662
 QY 507 ProHisTyrThrGlnAlaTrpSerGluAlaValPheIleHisGluGlyTyrThrHisAsp 526
 Db 1663 AATCACCCTCCCGC-----AGGTCAGGGCTCCACCCGGAGTACCTCAGAT 1710
 QY 527 Ala-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAsnLysValValIle 544
 Db 1711 GAGTCTACAAATTTTGAAGGGGAGATCGCCCTGTGAGAGCTGGAATAATGATCACCCTG 1770
 QY 545 AsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThr 564
 Db 1771 GGTCCCAACCTCCCTCCCATCTGCTCCCT- -GACACAGATACCTTTACGACCTG 1824
 QY 565 AspAspIleGlyThrAlaSerGlyTyrGlyLeuThrGlnArgGlyPheLeuAlaArgAsn 584
 Db 1825 GCGTGTATGGCTATGTCACTGCTGCGGGGTGATGAGAGAGAG---ATGCTCATGAC 1881
 QY 585 LeuMetTyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLys 604
 Db 1882 CTGAGTGTCTCGCTGCGCCGTAGCTAATCCACAGGCTGT----- 1923
 QY 605 ProProTyrProArgGlySer-----ValThrAlaAsnMetLeuCysAla 619
 Db 1924 GAGAACTGGCTCGGGGGAAGAAATAGAGATGATGTCTCTCAAAACATGTGTCTGCT 1983
 QY 620 GlyLeuGluSerGlyGlyLysAspSerCysArgGlyAspSerGlyLysAlaLeuValPhe 639
 Db 1984 GGAACCCCATCTCTAAAGCAGAGACGCTGCCAGAGGGGATATGTGGGCGCTTTTGGACGA 2043
 QY 640 LeuAspSerGluThrGluArgTrpPheValGlyLysIleValSerTrpGlySerMetAsn 659
 Db 2044 AGCGACCCGAACACATGATCGCTGGGCCACGCGCATGCTGCTCGGGGC---ATCGGG 2100
 QY 660 CysGlyGluAlaGlyGlnTyrGlyValTyrThrLysValIleAsnTyrIleProTyrPhe 679
 Db 2101 TGCAGCAGGGGC-----TATGGCTTCTACACCAAGGTGCTCAACTAGCTGACGTGATC 2154
 QY 680 Glu 680
 Db 2155 AAG 2157
 Db 2155 AAG 2157
 RESULT 7
 ABN95758 standard; DNA; 2493 BP.
 ABN95758:
 13-AUG-2002 (first entry)
 Gene #2256 used to diagnose liver cancer.
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 OS Homo sapiens.
 XX
 PN MO200229103-A2.
 XX
 PD 11-APR-2002.
 XX

PF 02-OCT-2001: 2001WO-US30589.
 XX 02-OCT-2000: 2000US-237054P.
 PR (GENE-) GENE LOGIC INC.
 XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JS;
 PI MPI: 2002-426119/45.
 XX
 DR Diagnosing and detecting the progression of liver cancer,
 XX hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample -
 PS
 PS Claim 1: SEQ ID NO 2256; 298bp; English.

CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytosolic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/publ/sequences.

CC Sequence 2493 BP: 619 A: 680 C: 662 G: 532 T: 0 other:

Alignment Scores:

Pred. No.:	8.77e-106	Length:	2493
Score:	1412.00	Matches:	301
Percent Similarity:	56.73%	Conservative:	108
Best Local Similarity:	41.75%	Mismatches:	248
Query Match:	37.57%	Indels:	64
DB:	24	Gaps:	22

US-09-874-198-2 (1-686) x ABN95758 (1-2493)

OY 1 MetArgLeuLeuThrLeuLeu-----GlyLeuLeuCys-----GlySerValAla 15
 64 ATGTGGCTCTTGTACTCTCTGTCGCCGCCCTGTTCTGACAGGCGACAGAGGCTCCATT--- 120
 DB 121 -----CCCATC-----CCTCAGAAAGTTATTGGGGAGGTGACTCCCTCGTTC 165
 OY 36 ProGlyGlyTyrAlaAsnAspGlnGluArgGlyTrpThrLeuThrAlaProGlyTyr 55
 166 CCCAACCCCTTACCCCACTTGAACACACCTGTGATCTACACTCCCAAGGATAC 225
 DB 166 CCCAACCCCTTACCCCACTTGAACACACCTGTGATCTACACTCCCAAGGATAC 225
 OY 56 ArgLeuArgLeuTyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGlnGluSerThr 75
 226 AGGGTAAACCTGCTCTTCCAGCAGCTTGAACCTTCTGAAAGCTGCTTCTATGAT 285
 DB 226 AGGGTAAACCTGCTCTTCCAGCAGCTTGAACCTTCTGAAAGCTGCTTCTATGAT 285
 OY 76 PheValHisLeuSerSerGlyAlaValValLeuAlaThrLeuGlyGlnGluSerThr 95
 286 TATGTCACAGATCTCTGCTATGAAGAAAGCCTGGGAGGTCTGTCGGCAACCTGGTCT 345
 DB 286 TATGTCACAGATCTCTGCTATGAAGAAAGCCTGGGAGGTCTGTCGGCAACCTGGTCT 345
 OY 96 AspThrGluArgAlaProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIle 115
 346 CCACTGGGCAACCCCGCCGGAAGAAAGAAATTTATGTCCTCCAGGGAACAAGATGCTGTC 405
 DB 346 CCACTGGGCAACCCCGCCGGAAGAAAGAAATTTATGTCCTCCAGGGAACAAGATGCTGTC 405
 OY 116 ThrPheArgSerAspTyrSerAsnGluLysPro-----PheThrGlyPhe 130
 406 ACCTTCACACAGACTTCTCCAAAGAGGAATGGACCATCATGTTCTTACAAAGGCTTC 465
 DB 406 ACCTTCACACAGACTTCTCCAAAGAGGAATGGACCATCATGTTCTTACAAAGGCTTC 465

OY 131 GluAlaPheTyrAlaAlaGluAspIleAspGluCys-----GluValAlaProGlyGlu 148
 466 CTGGCCCTACTACCAAGCTGTGGACCTTGATGATGTGCTTCCGGAGCAATATGAGGGAG 525
 DB 466 CTGGCCCTACTACCAAGCTGTGGACCTTGATGATGTGCTTCCGGAGCAATATGAGGGAG 525
 OY 149 -----AlaProThrCysAspHisHisCysHisAsnHisLeuGlyGlyPheTyrCys 165
 526 GAGATGCCCCAGCCCGCAGTGGCCAGACCTGTGTACACACTGTGGAGGCTACTCTGTG 585
 DB 526 GAGATGCCCCAGCCCGCAGTGGCCAGACCTGTGTACACACTGTGGAGGCTACTCTGTG 585
 OY 166 SerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSer 185
 586 TCCGTCCGTCCAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645
 DB 586 TCCGTCCGTCCAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645
 OY 186 GlyGluValPheThrGlnArgSerGlyGluLeuSerSerProGlyTyrProArgProTyr 205
 646 AGCGACTGTACACGAGGAGCATGAGCTATGCTCAGCTGGAGTACCTCGCTGCTTAC 705
 DB 646 AGCGACTGTACACGAGGAGCATGAGCTATGCTCAGCTGGAGTACCTCGCTGCTTAC 705
 OY 206 ProLysLeuSerSerCysThrTyrSerIleSerLeuGluGluGlyPheSerValIleLeu 225
 706 CCCCTGACCTCGCGCTGCAACTACACATCCGGGTGAGCGGGGCTTCAACCTGCACTTC 765
 DB 706 CCCCTGACCTCGCGCTGCAACTACACATCCGGGTGAGCGGGGCTTCAACCTGCACTTC 765
 OY 226 AspPheValGluSerPheAspValGluThrHisProGluThrLeuGlyProTyrAspPhe 245
 766 AAGTCTCTGGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 825
 DB 766 AAGTCTCTGGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 825
 OY 246 LeuLysIleGlnThrAspArgGluGluHisGlyProPheCysGlyLysThrLeuProHis 265
 826 CTACACATCATGCCAAGGAGAAACATTGGGCACTTCTGTGGAAAGCAAGGCCCC 885
 DB 826 CTACACATCATGCCAAGGAGAAACATTGGGCACTTCTGTGGAAAGCAAGGCCCC 885
 OY 266 ArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGluSerGlyAsp 285
 886 GACCTCGACACACGACGACAAATGCTGTGATCTGCTGTCTTTCACAGATGATGATGATGAT 945
 DB 886 GACCTCGACACACGACGACAAATGCTGTGATCTGCTGTCTTTCACAGATGATGATGATGAT 945
 OY 286 HisThrGlyTyrPylsIleHisTyrThrSerThrAlaGlnProCysProTyrProMetAla 305
 946 AGCCGGGCTGGAGCTGCGCTACACCGAGATCATCAAGTCCGCCCAAGACACC 1005
 DB 946 AGCCGGGCTGGAGCTGCGCTACACCGAGATCATCAAGTCCGCCCAAGACACC 1005
 OY 306 ProProAsn---GlyHisValSerProValGlnAlaLysTyrIleLeuLysAspSerPhe 324
 1006 CTAGAGAGATTCACATCATCCAGAACCTGACAGCTCAGTACAGATTCCTGACTACTTC 1065
 DB 1006 CTAGAGAGATTCACATCATCCAGAACCTGACAGCTCAGTACAGATTCCTGACTACTTC 1065
 OY 325 SerIlePheCysGluThrGlyTyrGluLeuGluGlnGlyHisLeuProLeuLysSerPhe 344
 1066 ATTGACTCTGCAAGCAGGCTACAGCTCATGAGAGGAGGACGAGTGTGCTGATCTTTC 1125
 DB 1066 ATTGACTCTGCAAGCAGGCTACAGCTCATGAGAGGAGGACGAGTGTGCTGATCTTTC 1125
 OY 345 ThrAlaValCysGlnLysAspGlySerTrpAspArgProMetProLacysSerIleVal 364
 1126 ACAGCTGTGCGCAGAGATGATGGACGTGACATGTCATGCTCCAGATCCAGATCAAG 1185
 DB 1126 ACAGCTGTGCGCAGAGATGATGGACGTGACATGTCATGCTCCAGATCCAGATCAAG 1185
 OY 365 AspCysGlyProProAspAspLeuProSerGlyArgValGluTyrIleThrGlyProGly 384
 1186 GACTGTGGGAGCGCCGGAACCTGCTATGATGATGATGATGATGATGATGATGATGATGAT 1245
 DB 1186 GACTGTGGGAGCGCCGGAACCTGCTATGATGATGATGATGATGATGATGATGATGATGAT 1245
 OY 385 ValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMetLys 404
 1246 GTGAACACTACAGAGCCGCTATGCTACCTGATGATGATGATGATGATGATGATGATGAT 1305
 DB 1246 GTGAACACTACAGAGCCGCTATGCTACCTGATGATGATGATGATGATGATGATGATGAT 1305
 OY 405 Val-----AsnAspGlyLysTyrValCysGluLysAlaAspGlyPhe 417
 1306 ACCAGAGCTGGCAGCAGGAGTCTGAGCAAGAGGCTGTACACCTGACAGCAGGACCAT 1365
 DB 1306 ACCAGAGCTGGCAGCAGGAGTCTGAGCAAGAGGCTGTACACCTGACAGCAGGACCAT 1365
 OY 418 Trp---ThrSerSerLysGlyLysSerLeuProValCysGluProValCysGly--- 435
 1366 TGGAAACATGACAGAGAGAGAGAG---ATTCTCTCGGTGCTTGGCAGGTGTGGAGAG 1422
 DB 1366 TGGAAACATGACAGAGAGAGAGAG---ATTCTCTCGGTGCTTGGCAGGTGTGGAGAG 1422
 OY 436 ---LeuSerAlaArgThrThrGlyGlyArgIleTyrGlyGlyGlnLysAlaLysProGly 454
 1423 CCGGTGAACCCCTGGAACAGAGGAGCCCATATGATGAGAGGCAAAAGCAAGATGAGGC 1482
 DB 1423 CCGGTGAACCCCTGGAACAGAGGAGCCCATATGATGAGAGGCAAAAGCAAGATGAGGC 1482
 OY 455 AspPheProGluValLeuIleLeuGlyGlyThrThrAlaAlaLeuLeuTyr 474
 1483 AACTTCCCTCGGAGGATGTTCCACCAACATCCAGGCGCGGGGCGGCGCTGCGGC 1542
 DB 1483 AACTTCCCTCGGAGGATGTTCCACCAACATCCAGGCGCGGGGCGGCGCTGCGGC 1542
 OY 475 AspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysAspAla----- 492

Db 656 GGTGGTACTTCCTGCTCCGCCCGGGAATATTCCTCCATGATGACATGAAGAAATTCG 715
 QY 181 SerAlaLeucysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGlu 200
 Db 716 GGGATTAATTTGGCGGGGATGATTCACCTGCACCTGATTTGGGAGATTCGAATCCCAAT 775
 QY 201 TyrProArgProTyrProIleuSerSerCysThrTyrSerIleSerLeuGluGluGly 220
 Db 776 TATCCCAACACATATCCAGAGAACTCAAGGTGATGATACAGATCCGGTTGGAGAAAGG 835
 QY 221 PheSerValIleLeuAspPhe---ValGluSerPheAspValGluThrHisProGluThr 239
 Db 836 TTCCAAGTGGTGGACCTTTCGCGAGAGAAAGATTTGATGTGAACACCTGACTCAGCG 895
 QY 240 LeucysProTyrAspPheLeuLysIleGlnThrAspArgGluGluHisGlyProPheCys 259
 Db 896 GGAAACTGCCCTTGACACTTTAGTTTGTTCGAGAGATCGCAATTTGGCTTCTACTGT 955
 QY 260 GlyLysThrLeuProHis-----ArgIleGluThrLysSerAsnThrValThrIleThr 277
 Db 956 GGCATGATGATTCCTCGGCTTAATATTTGAACCAAGATGATGCTGTGATATCATC 1015
 QY 278 PheValThrAspGluSerGlyAspHisThrGlyTyrPylsIleHisTyrThrSerThrAla 297
 Db 1016 TTCCAACATGATCTACACGGGCAAAAAGGCTGGAACCTTCGCTATCATGAGATCCA 1075
 QY 298 GlnProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLys 317
 Db 1076 ATGCCCTGCCCT---AAGGAAGACACTCCCAATTTCTGTTGGAGACCTCGCAAGGCAAAA 1132
 QY 318 TyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuGluGly 337
 Db 1133 TATGCTTTAGAGATGTGGCGAGATACCTGCTGATGCGTTGAAGTTTGAGAGGA 1192
 QY 338 HisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTyrPaspArgPro 357
 Db 1193 CGTGTGGTGCACATCTTTCTATTCGACTTGTCAAGCAATGGAAGTGGACTAATTC 1252
 QY 358 MetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgVal 377
 Db 1253 AATCGAATGTCAACCTGTGGAGCTGTGCATTCCTGATCCATTCGATGAATGTAAAGTT 1312
 QY 378 GluTyrIleThrGlyProGlyValThrTyrLysAlaValIleGlnTyrSerCysGlu 397
 Db 1313 GAA-----GACCCAGAGAGACACTTGTGTTGTTCTGTCATCGGCTACACTGTGAG 1363
 QY 398 GluThrPheTyrThrMetLys---ValAsnAspGlyLysSerTyrValCysGluAlaAspGly 416
 Db 364 GAGCCATATTAATACATGGAATAATGAGAGAGTGGGAGATATCATCTGCTGTGAACGGG 1423
 QY 417 PheTyrThrSerSerLysGlyGluLysSerLeuProValCysGluProValCysGlyLeu 436
 Db 1424 AGCTGGTGAATGAGTGTGGCGGCGGAGCTGCCAATATGTTTCCAGTCTTGAGATC 1483
 QY 437 SerAlaArg-----ThrThrGlyGlyArgIleTyrGlyGlyLysLysAlaLysProGly 454
 Db 1484 CCCAGAGAACCTTTGAAAAAAGAGATGATGAGAGATCCGATGAGATATTATAA 1543
 QY 455 AspPheProTyrGlnValLeuIleLeuGlyGlyThrThrAlaIleAlaLeuLeuTyr 474
 Db 1544 AACTTCCCTCGGCAAGTCT---TTCTTTCACAAACCCATGGGCTGGAGGCGCTATTAAAT 1600
 QY 475 AspAsnTyrValLeuThrAlaIleHisAlaValTyrGluGlnLysHis----- 490
 Db 1601 GAGTACTGGGTGCTGCGGCTGCTCATGTTGTGAGGGAACAGGAGCCACAAATGTAT 1660
 QY 491 -----AspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHis 508
 Db 1661 GTTGGGTCCACCTCACTGAGACCTCAGCGGTGGCAAAATCCAAAGATGCTCACTCT 1717
 QY 509 TyrThrGlnAlaTyrSerGlyAlaValPheIleHisGlyLysTyr----- 523
 Db 1718 -----GAGCATGTGTTTATTTCATCCGGGATGGAAGCTGTGGAAGTC 1759

QY 524 ---ThrHisAspAlaGlyPheAspAsnAspIleAlaIleLeuLysLeuAsnAlnLysVal 542
 Db 1760 CCAGAAAGACGACCAACATTTTGAATGATGACATTGTCAGCTGGTGGCGCTGAAGAACCCAGTG 1819
 QY 543 ValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMet 562
 Db 1820 AAATGAGGACCCACCTCTCTCCATCTGCTTACAGGACCTCTCTCCGACTCAACCTC 1879
 QY 563 ArgThrAspAspIleGlyThrAlaSerGlyTyrPylsLeuThrGlnArgGlyPheLeuAla 582
 Db 1880 ATGATGGGACCTGGGACTGATCTGAGCTGGGCGGAACAGAAAGAGATCTGTCT 1939
 QY 583 ArgAsnMetTyrValAspIleProIleValAspHisGlnLysCys---ThrAlaAla 601
 Db 1940 GTTCGGCTCAAGCGGCGCAAGCTTACTGTAGCTTCCTTTAAGAAATGCCAAGAAGTCGAAA 1999
 QY 602 TyrGluLysProProTyrProArgGlySer-----ValThrAlaAsnMetLeuCysAla 619
 Db 2000 GTGAGAAACCCACAGACAGATGCAAGGCTATGTTTCTTCACTCAATGATCTGTGCT 2059
 QY 620 GlyLeuGluSerGlyLysAspSerCysArgGlyAspSerGlyAlaLeuValPhe 639
 Db 2060 GGAGGAGAGAG---GGCATGATAGCTGTAAAGGGGACAGTGTGGCGCTTGTCTGTA 2116
 QY 640 LeuAsp---SerGluThrGluArgTyrPheValGlyGlyIleValIleSerTyrGlySerMet 658
 Db 2117 CAGGATCCCATATACAAAGACCAATTCACGACCTGGCTGTGCTGGGCGCC--- 2173
 QY 659 AsnCysGlyGluAlaGlyGlnTyrGlyValTyrThrLysValIleAsnTyrIleProTyr 678
 Db 2174 -----CAGTGTGGAGCTTATGGGCTCTACACACGGGTAAAGAACTATGTTGACTGG 2224
 QY 679 Ile 679
 Db 2225 ATA 2227
 Db 2225 ATA 2227
 RESULT 9
 AAC77788
 ID AAC77788 standard; cDNA; 2725 BP.
 AC AAC77788;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated gene sequence SEQ ID NO:182.
 XX
 KW Human: cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
 KW antidiabetic; antiaschemic; antirheumatic; antilarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 KW vasotropic; antipsoriatic; antitumorogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO20005350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587533/55.

DR P-PSDB: AAB43579.

XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX

PS Claim 1; Page 758-759; 2352BP; English.

CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerable; immunomodulator;
CC antidiabetic; antihistaminic; antihemorrhagic; antihistaminic;
CC antinflammatory; antitumor; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC neoplastic; vasotrophic; antiproliferative and angiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of hematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

SQ Sequence 2725 BP; 769 A; 582 C; 691 G; 681 T; 2 other;

Alignment Scores:

Pred. No.:	1,71e-91	Length:	2725
Score:	1238.00	Matches:	268
Percent Similarity:	52.788	Conservative:	102
Best Local Similarity:	38.233	Mismatches:	283
Query Match:	32.94%	Indels:	48
DB:	21	Gaps:	18

US-09-874-198-2 (1-686) x AAC77788 (1-2725)

QY 1 MetArgLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
DB 220 AAGTGTGATGCTGCTGCTTTTCACCTTTTGCGATGGTTATCTGACGCT----- 270
QY 21 LysTrpProGluProValIleGlyArgLeuAlaSerProGlyLeuProGlyLysValAla 40
271 -----ACCATGATGGGAGATCCTGCTCCCTAATCTCCTCAGGCAATGCC 318
41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyLysArgLeuArgLeuTyr 60
DB 319 AGTGAGGTAGAGAAATCTTGGGACATAGAAAGTTCCGAGAGGATAGGATCCCTCCTAC 378
QY 61 PheThrHisPheAspLeuLeuSerHisLeuGlyLysValAspPheValLysLeuSer 80
DB 379 TTCACCACTGTCGATGCTGTCAGAACTGTCGATGTCACGACGAGATGATC 438
QY 81 SerGlnAlaLysValLeuAlaThrLeuGlyGlnGluSerHisPheThrGlnAlaGala 100
DB 439 TCAGGAGACATGAGAAAGGAGGCTGTGTGACAGAGAGAGCATCAATCCCACTCT 498
QY 101 ProGlyLysAspThrPheThrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
DB 499 CCAATGTGTGAGAGTTCACAGTCCCATCAACAACCTCCAGGTGATCTTAAAGTCAGAC 558
QY 121 TyrSerAsnGlnLysProPheThrGlyPheGlnAlaPheThrAlaAlaGluAspIleAsp 140
DB 559 TTTTCCAAATGAGAGCGTTTACGGGGTTGCTGCTACATGTTCCACAGCAATTAAT 618
QY 141 GluCysGlnValAlaProGlyGlnAlaProThrCysAspHisCysHisAsnHisLeu 160
DB 619 GAATGACACAGATTTTGTAT---GATGTCCCT---GTAGCCACTTGTGACAAATTTCAAT 672

QY 161 GlyIlePheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
DB 673 GGAGGTACTTCTGCTCCCGCCCGGAAATTTCCATGATGATGACATGAATATTC 732
QY 181 SerAlaLeuGlySerGlyGlnValPheThrGlnAlaArgSerGlyGluLeuSerProGlu 200
DB 733 GGAGTTAAATGAGTGGGATGATATTCATCTGCATGATGGGGAATGCAATGCCAAT 792
QY 201 TyrProArgProThrProLysLeuSerCysThrThrSerLeuGlnGly 220
DB 793 TATCCCAACCATATCCACAGAACTCAAGGTGATACCAATCCGTTGGAGAAAGG 852
QY 221 PheSerValIleLeuAspPhe---ValGluSerPheAspValGluThrHisProGluThr 239
DB 853 TTCCAAGTGTGGTACCTTGGGAGAGAAAGATTTGATGTGAGAACACTGACACGCG 912
QY 240 LeuCysProTyrAspPheLeuLysIleGlnThrAspArgGlnGluHisGlyProPheCys 259
DB 913 GGAACTGCTTGACAGTGTATTTTGTGACAGAGATCGCAATTTGGCTTACTGT 972
QY 260 GlyLysThrLeuProHis-----ArgIleGluThrLysSerAsnThrValThrIleThr 277
DB 973 GTTCATGATTCCTCTGGGCTCAATATTTGAACCAAGATATGCTTATATATC 1032
QY 278 PheValThrAspGlnSerGlyAspHisThrGlyTyrLysIleHisThrSerThrAla 297
DB 1033 TTCCAACATGATCTACAGGCAAAAGGCGTGAACCTTCGATATCATGAGATCCA 1092
QY 298 GlnProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLys 317
DB 1093 ATGCCCTGCCCT---AAGGAAGACACTCCCAATTTCTTTGGAGCGCTCGAAGGAAAA 1149
QY 318 TyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGly 337
DB 1150 TATGCTTTAGAGATGTGGTGCAGATTAACCTGTCTGATGATTTCAATTTGGAGGGA 1209
QY 338 HisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerThrAspArgPro 357
DB 1210 CGTGTGTGATGCAACATCTTTTATTCGACTGTCAAGCAATGAAAGAGGATTAATCC 1269
QY 358 MetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgVal 377
DB 1270 AACTGAAATGTCACACCTGTGACCTGTGCAATTCGATTCATGAAATGTAAGTT 1329
QY 378 GluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnThrSerCysGlu 397
DB 1330 GAA-----GACCAAGAGACACTTTGTTGTTGCTGATCCGCTACACTTGTGAG 1380
QY 398 GluThrPheTyrThrMetLys---ValAsnAspGlyLysTyrValCysGlnAlaAspGly 416
DB 1381 GAGCCATTAATACATGAAATGAGAGAGGAGGAGATTCACACTGTCTGTAAACGG 1440
QY 417 PheThrThrSerSerLysGlyLysSerLeuProValCysGlnProValAlaGlyLeu 436
DB 1441 AGCTGGTGAATGATGCTGTGGCCGCGAGCTGCCAAATGCTTCACTGTGTGAGATC 1500
QY 437 SerAlaArg-----ThrThrGlyGlyArgIleTyrGlyGlnLysAlaLysProGly 454
DB 1501 CCCAGAGAACCTTTGAAACAAAACAGAGAGATTAATGAGAGATCCGATCGAGATATAA 1560
QY 455 AspPheProThrGlnValLeuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeuLeuTyr 474
DB 1561 AACTTCCCTGCGAATGCT---TTCTTTGCAACCCATGGGCTGGAGACCGCTCATTAAT 1617
QY 475 AspAsnThrValLeuThrAlaHisAlaValTyrGlnGlnLysHis----- 490
DB 1618 GATAGTGGTGTGACGCTGCTCATGTTGTGAGAGGAACAGAGGACCAACATGTAT 1677
QY 491 -----AspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHis 508
DB 1678 GTTGGGTCCACCTGATGAGACCTCAGCGCTGCGCAAAATCCAAAGATGCTCACTCCT 1734
QY 509 TyrThrGlnAlaThrPheSerGlnAlaValPheIleHisGlyLysTyr----- 523

[illegible]

Db	1657	ATTCGATGAGTACTGGGTCGTCGACGGCGGCTCAGCGTTGTGGAGGGAACCTGTGACCCAGT	1716
Qy	493	SerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGlnAla	512
Db	1717	ATGTATGTCCGGGTCCACACTTCTGAAATAAGACGGTTGAGA---AATGCCACAGGCTC	1773
Qy	513	TrpSerGlnAlaValPheIleHisGluGlyTyrThrHisAsp-----	526
Db	1774	ATCCCTGAACGTGATTATTCATCCAGCTGGAAACAAGAGGACCACTGAAATACACGG	1833
Qy	527	AlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAspLysValValIleAsnSer	546
Db	1834	ACAAATTTTGACATGACATGATGGCCCTGGTGCAGCATCAAGACCCGTGAATATGGGACCC	1893
Qy	547	AsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThrAspAsp	566
Db	1894	ACTGTTGCCCCCATCTCCTCGCAGAAACCTTCTCAGACTACACACCCCTCAGAGGTTGAC	1953
Qy	567	IleGlyThrAlaSerGlyTyrProGlyLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMet	586
Db	1954	CTGGGGCTGATCTCTGGGTGGGGCCGACAGACAGATTAGAACCAATGTTATTCACACTCAGA	2013
Qy	587	TyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProPro	606
Db	2014	GGGGCGAAGTTACCCATTAACATCTTTAGAAAAATGCCACAGAGTGAAGTGAAGAACCCG	2073
Qy	607	TyrProArgGlySer-----ValThrAlaAsnMetLeuCysAlaGlyLeuGluSer	623
Db	2074	AAAGCAGGCTCAACGACGACTATGTTTTCACGTGACACACATGATCTGTGCTGGG-----CA	2127
Qy	624	GlyGlyLysAspSerCysArgGlyAspSerGlyGly-----AlaLeuValPheLeuAsp	641
Db	2128	AAGGTGTGGACACTGTGAAGGTGGACAGCGGGGCTTTGTGCTCTCCGCTCCCAAT	2187
Qy	642	SerGlnThrGluArgTrrPheValGlyGlyIleValSerTrrGlySerMetAsnCysGly	661
Db	2188	GTCAGAGACCCCAATTCATATGTGGCTGGCTGGTGTCTCTGGGGGAAA-----	2235
Qy	662	GluAlaGlyLntYrGlyValTyrThrLysValIleAsnTyrIleProTrrPle	679
Db	2236	AAGTGTGGGACCTATGGGATCTACACAAAGTAAAGAACTACGTGACCTGGATC	2289
RESULT 11			
xx	AA592433	standard; cDNA; 2358 BP.	
xx	AA592433:		
xx	NC		
xx	DT	13-FEB-2002 (first entry)	
xx	DE	DNA encoding novel human diagnostic protein #28237.	
xx	xx		
xx	KW	Human: chromosome mapping; gene mapping; gene therapy; forensic;	
xx	KM	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
xx	OS	Homo sapiens.	
xx	PN	WO200175067-A2.	
xx	PD	11-OCT-2001.	
xx	PF	30-MAR-2001; 2001WO-US08631.	
xx	PR	31-MAR-2000; 2000US-0540217.	
xx	PR	23-AUG-2000; 2000US-0649167.	
xx	PA	(HYSE-) HYSEQ INC.	
xx	PI	Drmanac RT, Liu C, Tang YT;	
xx	xx	WPI: 2001-639362/73.	
xx	DR	P-PSDB; ABG28246.	

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1: SEQ ID NO 28237; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at fftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2358 BP; 607 A; 564 C; 646 G; 541 T; 0 other;

XX Alignment Scores:

Pred. No.:	3,54e-79	Length:	2358
Score:	1086.50	Matches:	248
Percent Similarity:	48.508	Conservative:	92
Best Local Similarity:	35.384	Mismatches:	248
Query Match:	28.914	Indels:	113
DB:	23	Gaps:	19

US-09-874-198-2 (1-686) x AAS92433 (1-2358)

OY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuSerValAlaThrProLeuGlyPro 20
DB 487 ATGTGTCATGTCCTGCTTTCATCTTGGCATGGGTTATGCTGAGCT----- 537
21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyAla 40
538 -----ACCATGTAATGGGAGATCTGCTCCCTAATCATCTCAGGCATATCCC 585
OY 41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyIleuArgLeuAlaGlyLeuTyr 60
DB 586 AGTGAAGTAGAAGAAATCTTGGACATAGAAAGTTCTGAAAGGATGGGATTCACCTCTAC 645
OY 61 PheThrHisPheAspLeuLeuSerHisLeuGlyGlnIleuArgPheValAlaLeuSer 80
DB 646 TTCACCATCTGCATATGAGCTGTCAAGAACTGTGGCATATGACATCAGTCAGATATATC 705
OY 81 SerGlyAlaLeuValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAla 100
DB 706 TCAGGAGACACTGAAAGAGGAGCTCTGTGGACAGAGGAGCATTAACAATCCCACTCT 765
OY 101 ProGlyLysAspThrPheThrLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
DB 766 CCAATGTGTGAAGAGTTCCATGCTCCATACAAACTCCAGGTGATCTTAAAGTCAGAC 825
OY 121 TyrSerAsnGlnLysProPheThrGlyPheGluAlaPheThrAlaAlaGlnAspIleAsp 140
DB 826 TTTTCAATGAAAGAGCGTTTACGGGTTTGCATACATATGTGGCCACAGACATAAAT 885
OY 141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisIscysHisAsnHisLeu 160
DB 886 GAATGCACAGATTTTGTATA--GATGTCCCT--TGTAGCCACATTCGACAAATTCATT 939

OY 161 GlyIlePheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
DB 940 GGTGTACTTCTGCTCTCTGCCCCCGGAAATATTTCTCATATGATGAAAGAAATTCG 999
OY 181 SerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyLeuLeuSerProGlu 200
DB 1000 GGAATTAAATTCAGTGGGATGATATTCATGACATGATTCGAGATTCGAACTCCAAAT 1059
OY 201 TyrProArgProThrProLysLeuSerSerCysThrTyrSerIleSerLeuGlnGly 220
DB 1060 TATCCCAACACATATCTCAGAACTCAGCTGATACCATCCGATCCGGTTGGAAAGCG 1119
OY 221 PheSerValIleLeuAspPhe--ValGluSerPheAspValGluThrHisProGluThr 239
DB 1120 TTCCAAGTGGTGGACCTTGGGAGAGAAAGATTTGATGTGAGACAGCTACGCG 1179
OY 240 LeuCysProThrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCys 259
DB 1180 GGAACCTGCTGACAGTTTAAATTTTGTTCAGAGATTCGCAATTTGGTCTTACTGT 1239
OY 260 GlyLysThrLeuProHis-----ArgIleGluThrLysSerAsnThrValThrIleThr 277
DB 1240 GGTCAATGATTCCTGCGGCTCTAAATATTAAGAAACCAAGATGCTGTGATATCATC 1299
OY 278 PheValThrAspLeuSerGlyAspHisThrGlyThrLysIleHisThrSerThrAla 297
DB 1300 TTCCAACTATATTAACAGGCGCAAAAAGGCTGGAAACTTGCTAT----- 1347
OY 298 GlnProCysProThrProMetAlaProProAsnGlyHisValSerProValGlnAlaLys 317
DB 1348 -----CATGAGATCTCT----- 1359
OY 318 TyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyLeuLeuGlnGly 337
DB 1359 ----- 1359
OY 338 HisLeuProLeuLysSerPheThrAlaValCysGlnLysAspLysSerThrAspArgPro 357
DB 1359 ----- 1359
OY 358 MetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgVal 377
DB 1360 -----GTGACGTGTGCGCATTCCTGAATCCATTTGAGATGGTAAAGTT 1401
OY 378 GluThrIleThrGlyProGlyValThrThrLysAlaValIleGlnIleThrSerCysGlu 397
DB 1402 GAA-----GACCACAGAGACACTTGTGTGTTCTGTCAATCCGCTACACTGTGTGAG 1452
OY 398 GluThrPheThrThrMetLys---ValAsnAspGlyLysThrValCysGluAlaAspGly 416
DB 1453 GACCCATATTACTACATGAAATAATGAGAGAGTGGGAGATGATCTGTGTTACGGGG 1512
OY 417 PheThrPheSerSerLeuGlyGlnLysSerLeuProValCysGlnLysGlyLeu 436
DB 1513 AGTGGGTGAATAGAGTGTGGGCGCGGAGCTGCCAAATGTGTTCACATCTGTGAGATC 1572
OY 437 SerAlaArg-----ThrThrGlyGlyArgIleThrGlyGlnLysAlaLysProGly 454
DB 1573 CCCAGAGAACCTTTGAAGAAAAACAGAGATTAATGAGAGATCCGATCAGATTAATTA 1632
OY 455 AspPheProThrGlnValLeuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeuTyr 474
DB 1633 AACTTCCCTGGCAAGTC--TTCCTTGAACACCCATGGGCTGTGGAGGCTCATTAAT 1689
OY 475 AspAsnTrpValLeuThrAlaAlaHisAlaValIleTyrGluGlnLysHis----- 490
DB 1690 GATTAAGTGGTGTGACGCGTCTCATGTTGTGAGGGAAGGAGGACCAACATGTAT 1749
OY 491 -----AspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHis 508
DB 1750 GTTGGGTCCACCTCAGTGCAGACCTCAGCGCTGGCAAAATCAAAAGTGTGCTCACTCT 1806

```

QY 509 TyrThrGlnAlaTrpSerGlnAlaValPheIleHisGlnGlyTyr----- 523
DB 1807 -----GAGCATGTGTTATTATTCACCGGATGAGTGGTGTGAAGTC 1848
QY 524 ---ThHisAspIleGlyPheAspAsnSPIleAlaIleuIleIysLeuAsnAnIysVal 542
DB 1849 CCAGAGAGGAGCAACATTGTAATGACATTGACATGCTGGCGCGGAGAACCCAGTG 1908
QY 543 ValIleAsnSerAnIleThrProIleCysLeuProArgSerGlnAlaGlnSerPheMet 562
DB 1909 AAATGAGACCCACCGCTCTCCATCTGCTACAGACACCTTCCGACTACACCTC 1968
QY 563 ArgThrAspSPIleGlyThrAlaSerGlyTyrPglyLeuThrGlnArgGlyPheLeuAla 582
DB 1969 ATGATGATGGGACCTGGACATCTCAGGCTGGGCGGAGAGAGAGATCGTCT 2028
QY 583 ArgAsnLeuMetTyrValAspIleProIleValAspHisGlnIysCys---ThrAlaIa 601
DB 2029 GTTCGCCTCAGGCGGAGGAGGTTACCTGTGCTCTTAAAGAAATGCCAAGAGTGAAA 2088
QY 602 TyrGlnLysProPtyrProArgGlySer-----ValThrAlaAsnMetLeuCysAla 619
DB 089 GTGAGAGAACCCACAGAGATGAGAGGCGCTATGTTTCATCTCTACATGATCTGTGCT 2148
QY 620 GlyLeuGlnSerGlyIysAspSerCysArgGlyAspSerGlyIysAlaLeuValPhe 639
DB 2149 GGAGGAGAGAGAG---GGCATGATAGCTGTAAAGCGGACGATGGGCGCTTGTCTGT 2205
QY 640 LeuAsp---SerIleThrGlnArgTrpPheValGlyGlyIleValSerTrpGlySerMet 658
DB 2206 CAGGATCCCATACACAGACCAATTCCTACGACGCTGGCTGTCTGGGGCC--- 2262
QY 659 AsnCysGlyGlnAlaGlyIleTyrGlyValTyrThrLysValIleAsnTyrIleProTyr 678
DB 2263 -----CAGTGTGGGACTATGGGCTCTACACAGCGGTAAAGAACTATGTGACTGG 2313
QY 679 Ile 679
DB 2314 ATA 2316
RESULT 12
AAF18150
ID AAF18150 standard; DNM; 2063 BP.
AC AAF18150;
XX 14-MAR-2001 (first entry)
DT Lung cancer associated polynucleotide sequence SEQ ID 169.
XX
XX Human: lung cancer associated protein; neuroprotective; cytoskeletal;
XX cardioactive; immunomodulatory; muscular active; vulnery;
XX gastrointestinal; nephrotoxic; antineoplastic; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
XX OS
XX PN MO20005180-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000MO-US05918.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX WPI: 2000-587514/55.
XX P-PSDB: AAB58274.

```

```

XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -
XX
XX Claim 1: Page 632-633; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective, cytoskeletal, cardioactive,
CC immunomodulatory, muscular active general; vulnery; gastrointestinal
CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
XX Sequence 2063 BP; 544 A; 568 C; 515 G; 431 T; 5 other;
SQ
Alignment Scores:
Pred. No.: 8, 82e-62 Length: 2063
Score: 873.00 Matches: 201
Percent Similarity: 46.39% Conserved: 69
Best Local Similarity: 34.54% Mismatches: 167
Query Match: 23.23% Indels: 145
DB: Gaps: 16
US-09-874-198-2 (1-686) x AAF18150 (1-2063)
QY 227 PheValGlnSerPheAspValGlnThrHisProGlnThrLeuCysProTyrAspPheLeu 246
DB 3 TTCCTGGAGCGCTTTATATGATGACACACAGCAANTACAGCGCCATGACGACGTA 62
QY 247 LysIleGln----- 249
DB 63 CAGGTACAGCGCTCTACCTGAAGAGCCCTCCCTCTGTGTGCACTTGCTCT 122
QY 249 ----- 249
DB 123 CTGTGCCAATCTCTCTGATCCCTGGCCAGCTGGGGCAGAGAACGCAATCACC 182
QY 249 ----- 249
DB 183 CATGGGCTGGGTAAAGTCCACACACACTAGAAATGGGTGATGGGATCCCTTTACATT 242
QY 249 ----- 249
DB 243 CCCCTGAAAAACACATAAGGAGGAGATTTCATCACCACACACACACCTGGCCACAG 302
QY 249 ----- 249
DB 303 CTACTACCAAGTTGGCGCTGTGTAAGAACGTCACAGTGAAGAAAAAACCTCTCTCC 362
QY 250 -----ThrasPArgGlnGlnHis-GlyProPheCysGlyLysThrLeuProH 265
DB 363 CTACTAGATCTATGCCAAGGAGAGAACATTTGGCGAGTCTGTGGGAGACAAAGCTCC 422
QY 265 sarGlnGlnThrLysSerAsnThrValThrIleThrPheValThrAspGlnSerGlyAs 285
DB 423 CGACCTCGACACAGCAGCAATGCTGTGATCTGCTCTTCAACAGATGAGAGCGGGGA 482
QY 285 pHisThrGlyTyrPlyIleHisTyrThrSerThrAlaGlnProCysProTyrProMetAl 305
DB 483 CAGCGGGGCTGGAAAGCTGCGCTACACACAGAGATCATCAAGTGGCCCGGAGCCAAAGAC 542

```

Oy	305	aproProAsn----	GlyHisValSerProValGlnAlaIleuTyrrTleuLysAspSerPh	324	
Db	543	CTTgAGAGGAGTTCACCATCATCCAGAACTGGACGCTCATGTACCACTTCGGTACTT		602	
Oy	324	eSerTlePhecysGluThrGlyTyrGluLeuLeuGlnGlyHisLeuProLeuLysSerPh	344		
Db	603	CATTGTACCTGCAGAACAGAACCGCTTACCAAGCTCATATAGGGGGAACAGAGTGTGCATTCTT		662	
Oy	344	eThrAlaValCysGlnLysAspGlySerThrAspArgProMetProAlaCysSerTleVa	364		
Db	663	CACAGCTGTGTGGCAGAGTATATGGACAGCGGATCGTCCAGTCCCAATCATCAGATCAA		722	
Oy	364	IASpCysGlyProProAspAspLeuProSerGlyArgValGluTyrTleThrGlyProGly	384		
Db	723	GGACTGTGGGAGACCGCGAAGAACTGGCTATAGTGACTTCCTTACACACCAATGG		782	
Oy	384	yValThrThyTyrLysAlaValIleGlnTyrSerCysGluGlnIuThrPheTyrThrMetLys	404		
Db	783	AGTAAACCTACAGAGCCCGTATCCAGTACTACCTCCATGGCCATTTTCAAGATCA		842	
Oy	404	sVal-----	AsnAspGlyLysTyrValCysGluAlaAspGlyPh	417	
Db	843	GACCAGAGCTGGCAGCAGGAGTCTAGCAAGGGGTGTACACTCCACAGCAGCAGGGCAT		902	
Oy	417	eTrrP-----	ThrSerTleGlyGlyLysSerLeuProValCysGluProValCysGly--	435	
Db	903	TTGGAAATATGAACAGAGAGGAGAGAAAG--	ATTCTCGTGTCTTGGCAGTGTGGGAA	959	
Oy	436	-----	LeuSerAlaArgThrThrGlyLysArgTyleTyrGlyGlnLysAlaLysProGly	454	
Db	960	GCCCCGATAACCCCGTGGACAGACAGGACGAGCATGGAGGGCAAAAAGCCAAATATGG		1019	
Oy	454	yAspPheProTrrPglInValLeuIleuGlnGlyThrThrAlaAlaGlyAlaLeuTyr	474		
Db	1020	CAACTTCCTCCCTGGGAGGTGTTCACAAATTCACAGGGGGCGGGGGCGGCTGTCTGG		1079	
Oy	474	rAspAsnTrrValLeuThrAlaAlaHisAlaValTyrGlnGlnLysHisAspAla----	492		
Db	1080	CGACCGGTGATCTCTCACACTTCCACACACCTGTATCCAAAGGAACAGAAAGCCAAAG		1139	
Oy	493	-----	SerAlaAspIleArgTyrMetGly-----	ThrLeuLysArgLeuSe	506
Db	1140	CAACGCTCTTTGGATGTGTCTCTGGGCCACACAAATGTGGAAAGCTCATGAAGCTTAGG		1199	
Oy	506	rProHisTrrThrGlnAlaIleTrrSerGluAlaValPheIleHisGlnGlyTyrThrHisAs	526		
Db	1200	AAATCACCCCATCCCGCAGGNTCAGC-----	GTCCACCCGGACTACCGTCAAGA	1247	
Db	526	PalA-----	GlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValAlaI	544	
Oy	544	eAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGlnSerPheMetArgTh	564		
Db	1308	GGGTCGCAACCTCTCTCCCATTCGCTCTCTCT-----	GACAAAGTACTCTTACAGACT	1361	
Oy	564	rAspAspIleGlyThrAlaSerGlyTrrPclLysLeuThrGlnArgLysPheLeuAlaArgAs	584		
Db	1362	GGCGTTGATGGCTATGTCACTGGCTTCGGGGTCATAGGAGGAAG--	ATTGCTCATGA	1418	
Oy	584	nLeuMetTrrValAspIleProIleValAspHisGlnLysCysThrAlaIleTyrGlyLys	604		
Db	1419	CTGTAGTGTTCCTCGCTGGCTGGCTACTTAATCACAAGCCTCT-----		1461	
Oy	604	sProIrrTrrArgGlySer-----	ValThrAlaAsnMetLeuCysAl	619	
Db	1462	GAGAACTGGCTCCGGGGAAGAAATAGATGATGATGTGTCTCTCAAAACATGTTCTGTGC		1520	
Oy	619	agLysLeuGlnSerGlyLysLysAspSerCysArgGlyAspSerGlyGlyAlaLeuValPh	639		
Db	1521	TGGACACCCATCTTAAAGGAGGACCTCTGCACAGGGGCAATGTGGGGCGTTTGGCAGT		1580	

Oy	639	eluaspeerclnhrclgnarqrrpphevalglgylilleva1setrpglysermevls	659
Db	1561	AAGGACCCCGAACACTATCGCTGGGCGCCAGGGCATGCTCTGGGGC---ATCGG	1637
Oy	659	ncysgclgclua1aol1gclntfrclyalvaltrythrlvsval1leasntry1leprotrp11	679
Db	1638	GTGCGACGAGGGGC-----TATGGCTTCTACACCAAGTGCCTCACTACTGAGACTGCAT	1691
Oy	679	eglu 680	
		:::	
Db	1692	CAAG 1695	
RESULT 13			
ID	AAFI8149		
XX	AAFI8149	standard; DNA: 1148 BP.	
AC	AAFI8149;		
XX			
DT	14-MAR-2001	(first entry)	
XX			
DE	Lung cancer associated polynucleotide sequence SEQ ID 168.		
XX			
KW	Human; lung cancer associated protein; neuroprotective; cytoskeletal;		
KW	cardioactive; immunomodulatory; muscular active; vulnerrary;		
KW	gastrointestinal; nephrotoxic; antilinfective; gynecological;		
KW	antibacterial; diagnosis; neural disorder; immune disorder; reproductive		
KM	proliferative disorder; wound healing; infectious disease; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO20005180-A2.		
XX			
PD	21-SEP-2000.		
XX			
PF	08-MAR-2000; 2000WO-US05918.		
XX			
PR	12-MAR-1999; 99US-0124270.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(ROSE/) ROSEN C A.		
XX			
PI	Ruben SM;		
XX			
DR	WPI: 2000-587514/55.		
XX	P-PSDB: AAB56273.		
PT	Lung cancer associated gene sequences, referred to as lung cancer		
PT	antigens, useful for treatment, prevention, and diagnosis of disorders		
PI	such as lung cancer -		
XX			
PS	Claim 1; Page 632; 1425pp; English.		
XX			
CC	Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer		
CC	associated proteins represented in AAB58106 - AAB58548. Lung cancer		
CC	associated proteins and polynucleotide sequences, their agonists, and		
CC	antagonists may have neuroprotective; cytoskeletal; cardioactive;		
CC	immunomodulatory; muscular active general; vulnerrary; gastrointestinal		
CC	general; nephrotoxic; antilinfective; gynecological; or antibacterial		
CC	activity. The invention also includes antibodies specific for the		
CC	protein or polynucleotide sequences. The lung cancer associated		
CC	polynucleotide sequences may be used for detection of lung cancer,		
CC	chromosome identification, as chromosome markers, and for numerous other		
CC	diagnostic or research purposes. The proteins may be used to treat		
CC	disorders such as neural, immune, muscular, reproductive,		
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative		
CC	disorders. The proteins may also be used in the treatment of wounds and		
CC	infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and		
CC	peptide AAB5549 are used in the course of the invention for the		
CC	identification and characterisation of the polynucleotide and protein		
CC	sequences.		
XX			
SO	Sequence 1148 BP; 258 A; 341 C; 305 G; 242 T; 2 other;		

Alignment Scores:

Pred. No.:	4,29e+47	Length:	1148
Score:	689.50	Matches:	142
Percent Similarity:	57.798	Conservative:	51
Best local Similarity:	48.518	Mismatches:	119
Query Match:	18.358	Identities:	22
DB:	21	Gaps:	7

US-09-874-198-2 (1-686) x AAF18149 (1-1148)

QY	1	MetrArgLeuLeuThrIleLeu-----GlyLeuLeuCys-----GlySerAla	15
Db	76	ATtGGCTCTTtGACtCCtCGtGtCCGGCCtCTTtCGACAGGcCAGAGGCTtCATt---	132
QY	16	ThrProLeuGlyProLysrTyrrProGluProValPheGlyArGLeuAlaSerProGlyPhe	35
Db	133	---CCCATC-----CCtCAAGAtGtATTtGGGGAGGtGAtCTtCCtCTGtTC	177
QY	36	ProGlyGluTyrAlaAsnAspGlnGluAArgTrrPheThrAlaProProGlyTyr	55
Db	178	CCCAAGcCTtACCCCAACCACTTtGAAACACACtGTGTATCtACAGtCCtCCAGGATAC	237
QY	56	ArgLeuArGLeuTyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAsp	75
Db	238	AGGGtGAAGcCTGtCTtCCACAGAtTtGAtCTtGAGcCTtCGAAGcGtCTTATGAT	297
QY	76	PheValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuGlyGlyGlnGluSerThr	95
Db	298	TATGTCAAGAtCTtCTGCTGATtAGAAAGcCTtGGGAGGtGTCTGTGTGGCACTGGTCT	357
QY	96	AspThrGluArGAlaProGlyLysAspThrPheTyrSerLeuGlySerLeuAspIle	115
Db	358	CCACTGGGCAACCCcCCGGGAAAGAGGAATtTATGTCCCAAGGAAcAAGAtGTGTG	417
QY	116	ThrPheArGSerAspTyrSerAsnGlyLysPro-----PheThrGlyPhe	130
Db	418	ACCTtCCACACAGcCTtCTCCAAcGAGAGAtAGtAGGACATATGtTTCTTACAGGGCTTC	477
QY	131	GluAlaPheTyrAlaAlaGlnAspIleAspGluCys-----GlnAlaAlaProGlyGlu	148
Db	478	CTGGcCTACtACCAAGcGTGTGAACCTtGATGATtGTGCTtCCCGAGCAAAATCAGGGAG	537
QY	149	-----AlaProThrCysAspPheHisHisCysHisAsnHisLeuGlyLeuPheTyrCys	165
Db	538	GAGAGTCCCAcGAGCCcAGTCCcAGACcCTGTGTACACAGtACCTtGGAGGtACtTCTGT	597
QY	166	SerCysArGAlaGluTyrValLeuHisArgAsnLysArgTrrHisCysSerHisLeuGlySer	185
Db	598	TCTtGcCCGTCCAGGcCTtATGAGcCTtCAcGAAGAtAGGAtCTtCTCGcCAGcGTAGtGAC	657
QY	186	GlyLysValPheThrGlnArGSerGlyGluLeuSerSerProGluTyrProArgProTyr	205
Db	658	AGcGAGcGTACACAGcGAGGAGcGATACATCTCCAGcCTGTGAGtACCTtGGTCTTAC	717
QY	206	ProLysLeuSerSerCysThrTyrSerTLeSerLeuGlnGluGlyPheSerValIleLeu	225
Db	718	CCCCcTACcCTGGcCTGTCAcTACAGAtCTCGGtGGtGAGcGGGcCTCAcCTGTCAcCTC	777
QY	226	AspPheValGluSerPheAspValGluThrHisProGluThrLeuGlyProTyrAspPhe	245
Db	778	AAGTtCTCTGGAGcCTTtTGAATtATGAtGACACcACcAGAtACtATACtGcCCCTATGTCCAG	837
QY	246	LeuLysTLeGlnThrAspArGArgGlnGluHisGlyProPheCysGlyLysThrLeuProHis	265
Db	838	CTACAGAtCTATAGcCAACGGGAGAGAtCATtTGGGAGtTCTGTGTGGAAcGAAGcCCcCC	897
QY	266	ArgTLeGluThrLysSerAsnThrValThrIleThrPheValThrAspLysSerGlyAsp	285
Db	898	GACCTCGACACcACAGACcATGcTGTGATCTCTGtCTCTTCACAGAtGATGTcGGGGAGc	957
QY	286	HisThrLysThrLysLysIleHisTyrThrSerThrAlaGln-ProGlyProTyrProMetAl	305
Db	958	AGcCGGcGTGGAGcCTGcCGcCTACcACcAGcAGAtCATCAAtGTCCcCCcACcAGAGc	1017

```
QY 305 aProProaNsGLYHisValSerProValGlnAlaLysTyr 318
      ||| ||| ||| :::
Db 1018 CCTAGACGAGTTCACCATCATCCAGAACCTGCAGCCTTC 1057
```

RESULT 14
AAK53540/c
ID AAK53540 standard; cDNA; 371 BP.

AC AAK53540;

DT 16-NOV-2001 (first entry)

DE Murine transport and binding associated protein encoding cDNA SEQ ID 105.

KN Murine liver: gene library: amino acid synthesis: binding protein:
 KN cell metabolism: energy metabolism: fatty acid metabolism: synthesis:
 KN phospholipid metabolism: purine: pyrimidine: nucleoside: nucleotide:
 KN replication: transcription: translation: transport protein: ss.

OS Mus musculus.

PN DE20103510-U1.

PD 07-JUN-2001.

PF 28-FEB-2001; 2001DE-2003510.

PR 02-DEC-1999; 99DE-1058160.

PA (LION-) LION BIOSCIENCE AG.

DR WPI; 2001-368570/39.

PT Gene library containing sequences with specific 3'-ends and no polyA tail, encoding proteins involved in a wide range of cellular processes

PS Claim 15; Page 51-52; 251pp; German.

CC This invention describes a novel gene library (A) comprises a gene
CC sequence (or its part) encoding a protein involved in amino acid
CC synthesis, cellular/energy metabolism, metabolism of
CC fatty acids/phospholipids, synthesis or breakdown of
CC purines/pyrimidines/nucleosides/nucleotides, DNA
CC replication/transcription/translation, or is a transport/binding protein.
CC (A) are produced that correspond to the 3'-end of mRNA but without the
CC polyA tail. They can be prepared more efficiently and with less effort
CC than conventional libraries. AK53436-AK54275 represent fragments of the
CC gene library described in the method of the invention.

SQ Sequence 371 BP; 94 A; 104 C; 68 G; 105 T; 0 other;

Alignment Scores:

Pred. No.:	3,29e-35	Length:	978
Score:	536.50	Matches:	38
Percent Similarity:	89.92%	Conservative:	9
Best local Similarity:	82.35%	Mismatches:	11
Query Match:	14,28%	Indels:	1
DB:	22	Gaps:	1

US-09-874-198-2 (1-686) x AAK53540 (1-371)

QY	568	GLIHLAASRGYLPGLYLENTHGIAAGGLYPHEULALAVASLEUMLTYR	587
		
Db	365	GCAACTGTGGCTGGGGGTTAACCACCAAGGGCTTTGCTGAACCACTAAATTTT	306
		
QY	588	VALASPILEPROLLEVALASPHISGLINLYSCSYTHRIALALATRYGLULYSPROPTOTYR	607
		
Db	305	GTGGACATACCAATTCGTGACACCAAAATGTACCAACCGTGATGATAAAG---CTCTAT	249
		
QY	608	PROARGLSERVALTHRIALASMETLEUCYSLAAGLYLEUGLUSERYGLYGLYASP	627
		

Oy	26	raiaGlnProCysproCtyrTrpMetAlaProProAsnGlnHisValSerProValGlnAl	316
Db	450	-AGTCAGCCCATTCAGC-----	464
Oy	316	alysTryIleLeuLysAspSerPheSerIlePheCysGluThrGlyTYrGluLeuG1	336
Db	465	-----GAGCCCAAGCAG	475
Oy	336	ngLIHisIleuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAsp	356
Db	476	gggGCT-----	482
Oy	356	gProMetProAlaCysSerIleValAspCysGlyProProAspLeuProSerGlyAr	376
Db	482	-----	482
Oy	376	gValGluTYrIleThrGlyProGluValThrTYrLysAlaValIleGlnTYrSerCy	396
Db	483	-----GAGCCCTCAGCCACCTGCAGACACCTGCCAAG-----GTCAGAACCACTG	532
Db	396	sgLglnLutThrPheTYrThrMetLysValAsnAspGlyIstYrValCysGlnAlaSpG1	416
Oy	416	yPheTrPTrHiserSerIstGlyLysSerLeuProValCysGluProValCysGly--	435
Db	533	CGAGGAGCCCTATTAT--CAGGCCGGCGCAGCAGGGGCACTCACCTGTGCACCCACGG	589
Oy	590	GACCTGGAAAGACAGACAGAGTGGGGAGAGGTTCCTCAGTGAATGCCCTGTGGCAGG	649
Oy	436	-----LeuSerAlaArgThrGlnGlyIstYrGlyIstYrGlyGlnLys	450
Db	650	GCCAGTACCCCCCATTCGCCAGATACAGAGACC-----CTCCGCTTCACG	697
Oy	450	salAlstProGlyAspPheProTrpGlnValIleuIleGlnGlyTYrThrAlaAlG1	470
Db	698	AGCCAGCTGGGCAACTTCCCTGGCAGAGCCTCACACATGCCAGCGCGTGGGGGGG	757
Oy	470	yAlaLeuLeuTYrAspAsnTrpValLeuThrAlaAlaHisAlaValTYrGluGlnLys--	489
Db	758	GGCCCTGCTGGGGACAGATGGATCTCTCATCTGCTGCCACACCGTCTACCCAGAGCAG	817
Oy	490	-----HisAspLeuIleAlaLeuAspIleAlaGlyMetGly-----Th	501
Db	818	TGTTTCTCTCAGAAAGACAGAGTGGATGTCTTCTGGGCCACACAGCCATGATGA	877
Oy	501	rLeuLysArgLeuSerProHisTYrThrGlnAlaTrpSerGlnAlaValaPheIleHisG1	521
Db	878	GATCTGAAATGGGGAGAACCAACCCTGTC-----CACCGTGGTGTGCACCC	925
Db	521	wGlyTYrThrHisAspAla-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAs	539
Db	926	CGACTACCGTCAGATATAGTCCATTAATCTTACGGGGGACATCGCCCTCTGGACTCTCA	985
Oy	539	nasnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGlnAlaG1	559
Db	986	GCACAGCATCCCCCTGGGCCCAAGCTCCCGGCTGTCTGGCC-----GATTAAGA	1039
Oy	559	uSerPheMetArgThrAspAspIleGlyThrAlaSerGlyTYrPrgLysLeuThGlnArgG1	579
Db	1040	GACCTCTACCGACGGCGCTGTGGCTACGTCACTGCGTGGTGGGATAT--GAAATGGG	1096
Oy	579	yPheLeuAlaArgAsnLeuMetTYrValaLaspIleProIleValaLaspHisGlnLysCysTh	599
Db	1097	CTGGCTAACTACTAGCTGAAGTACTGCAAGGCTGCTTGAAGTCCCGAGGAGGCTGCA	1156
Oy	599	raIaAlaTYrGluLysProProTYrProArgLysSerValThrAlaAsnMetLeuCysAl	619
Db	1157	CGCCTGCTCCAAAGAGAGACAGAGACCCGAG--GTTTTCCTGACAAATATGTCTGTGT	1213
Oy	619	agLYLeuGlnLysrClyLysAspSerCysAlaGlyAspSerGlyGlyAlaLeuValPh	639
Db	1214	TGGGGATGAGACGCAAGAGCAGACAGTGTCTTCCAGGGGGAGACAGTGGCAGTCTATGTGT	1273

Qy	639	eLeaAspsercluhrcLnurgrtprhveAlGyglIleVslserTprglSerMeclAs	659
Db	1274	ATGGGACAATCATCCATCTACTGCGTGGCCAGGCGCATGTGCTCGGGCG--ATAGG	1330
Qy	659	nCySGlVcJuaIaGlyIntTyrGlyValTyrThrLysValIleAsnTyrIleProTprI1	679
Db	1331	GTGTGCGCCAAAGG-----TATGACTTCTACACCAAGGCTGCACGTATGTGACTGCAT	1384
Qy	679	eGluAsnIleIleSer 684	
Db	1385	CAAGGAGTGTGATGAAT 1400	
RESULT 16			
ID	AA514091	standard; DNA: 1667 BP.	
XX	AA514091:		
AC	AA514091:		
XX	18-DEC-2001	(first entry)	
DT			
XX			
DE	Human FcTR5a DNA sequence.		
XX			
KW	Human; FCPRX; myelogenous leukaemia; carcinoma; melanoma; glioma; ds;		
KW	astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;		
KW	neurological disorder; neurodegenerative disorders; nerve trauma;		
KW	familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;		
KW	demyelinating gardner syndrome; familial myelodysplastic syndrome;		
KW	mental health condition; immunological disorder; allergy; infertility;		
KW	bronchial asthma; Avellino type eosinophilia; lung disease; deafness;		
KW	reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;		
KW	desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;		
KW	gastric disorders; pancreatic disease; Schistosoma mansoni infection;		
KW	Spinocerebellar ataxia; Plasmodium falciparum paraisitemia; diabetes;		
KW	Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;		
KW	RetS-bucklers corneal dystrophy; cystostatic; immunosuppressive;		
KW	antiallergic; antisthmatic; antifertility; antiinflammatory;		
KW	antidiabetic; protozoicide; hepatotropic; virucide; ophthalmological;		
KW	gynaecological; antifertility; immunostimulant; auditory; haemostatic;		
KW	gene therapy; FcTR5a; human complement C1r component precursor.		
XX			
OS	Homo sapiens.		
XX			
FH		Location/Qualifiers	
FT	CDS	34..1497	
FT		/*tag= a	
FT		/product= "Human FcTR5a"	
FT	5'UTR	1..33	
FT		/*tag= b	
FT	3'UTR	1498..1667	
FT		/*tag= c	
XX			
PN	WO20016747-A2.		
PD	13-SEP-2001.		
XX			
PF	05-MAR-2001; 2001WO-US07160.		
XX			
PR	03-MAR-2000; 2000US-186592P.		
PR	03-MAR-2000; 2000US-186718P.		
PR	06-MAR-2000; 2000US-187293P.		
PR	06-MAR-2000; 2000US-187294P.		
PR	17-MAR-2000; 2000US-190400P.		
PR	07-APR-2000; 2000US-196018P.		
PR	03-JAN-2001; 2001US-259548P.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Vernet CAM, Fernandes E, Shinkets RA, Herrmann JL, Majumder K;		
PI	MacDougall J, Mishra V, Mezes PS, Rastelli L,		
XX	WPI: 2001-596837/67.		
DR	P-PSDB: AAU08683.		
XX			

PT Novel polypeptides designated as FCTR polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
PT states

PS Claim 9: Page 93-94: 215pp: English.

XX The invention relates to human FCTR polypeptides, FCTR1-FCTR7, and the
CC nucleic acids encoding them. These sequences are useful for the treatment
CC or prevention of numerous disorders including myelogenous leukaemia,
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
CC allimmune thrombocytopenia, neurological disorders, neurodegenerative
CC disorders, nerve trauma, familial myelodysplastic syndrome,
CC Charcot-Marie-Tooth neuropathy, demyelinating gardner syndrome, familial
CC myelodysplastic syndrome, mental health conditions, immunological
CC disorders, allergy and infection, bronchial asthma, Avellino type
CC eosinophilia, lung diseases, reproductive disorders, infertility, male
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
CC desmoid disease, tumor syndrome, liver cirrhosis, hepatitis C, gastric
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
CC infection, spinocerebellar ataxia, Plasmodium falciparum parasitaemia,
CC Corneal dystrophy-Greponow type I, Corneal dystrophy-lattice type I and
CC Reis-Bucklers corneal dystrophy. This sequence represents DNA encoding
CC FCTR5a, a protein bearing sequence homology to human complement C1R
CC component precursor.

XX Sequence 1667 BP: 354 A; 475 C; 498 G; 340 T; 0 other:

Alignment Scores:
Pred. No.: 4,756-31 Length: 1667
Score: 496.00 Matches: 137
Percent Similarity: 33.67% Conservative: 74
Best Local Similarity: 22.89% Mismatches: 194
Query Match: 13.20% Indels: 261
DB: 22 Gaps: 20

US-09-874-198-2 (1-686) x ASI14091 (1-1667)

QY 17 ProLeuGluProLys-TrpProGluProValPheGluAlaLeuSerProGluPheP 36
DB 159 CCTGTGGCCCAAGAGCTACCCAG-----CAGCTGACATCCCGCGGATACC 206
QY 36 oclYgluYTrAlaAsnAspGlnGluArgTrPThrLeuThrAlaProGluYTrAr 56
DB 207 AAGCCGCTGTGGCAAGAGCGAGAGCGACGACATCAAGCGCTGAGAGGCTTTGC 266
QY 56 gLeuArGLeuYrPheThrHisPheAspLeuGluLeuSerHisLeuGluYTrAsp 76
DB 267 TGTGAGGCTGCTTCCAGAGCTGAGAGCTGAGCGCTCCAGAGCTGTGCAAGGAGCTC 326
QY 76 eValYLeuSerSerGlyAlaLysValLeuAlaThrLeuGlyGlnGluSerThrAs 96
DB 327 TGTCAACAATCTATTCGTGGGTTCGATGCAAGCCAGTTCTGTGGTACAGAGCTCCCC 386
QY 96 pThrGluAlaGalaProGlyLysAspThrPheYrSerLeuGlySerSerLeuAspThr 116
DB 387 TCTGGGACAGCCCGCTGACAGAGAGCTTTATCTATCTACGAGAGAGATTGGCGCTGAC 446
QY 116 rPheArGSerAspYrSerAsnGluLysProPheThrGlyPheGluAlaPheYrTrAla 136
DB 447 CTTCCGACACAGCCCTCTCGGAG----- 471
QY 136 aGluAspIleAspGluGlyGlnAlaLalaProGlyGluAlaProThrCysAspHisCys 156
DB 471 ----- 471
QY 156 sHisAsnHisLeuGlyGlyPheYrCysSerCysArgAlaGlyYrValLeuHisArgAs 176
DB 472 -----AA 473
QY 176 nLysArgThrCysSerAlaLeuGlySerGlyGlnAlaPheThrGlnArgSerGlyGlu 196
DB 474 CAAG----- 477

QY 196 uSerSerProGluYrProArgProYrProLysLeuSerSerCysYrThrYrSerIleSe 216
DB 478 -----ACTGCCA 485
QY 216 rLeuGlnGluGlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrH 236
DB 486 CCTCCACAGGGGCTTCGTGGCCTTAC----- 513
QY 236 sProGluThrLeuGlyProYrAspPheLeuLysIleGlnThrAspArgGluLuhIsG 256
DB 513 ----- 513
QY 256 yPropheCysGlyYrThrLeuProHisArgIleGluThrYsSerAsnThrValThrI 276
DB 514 -----CAACCGTGGCTGT 527
QY 276 eThrPheValThrAspGluSerGlyAspHisThrGlyTrpLysIleHisYrThrSerTh 296
DB 528 GAACGTAT----- 534
QY 296 rAlaGlnProCysProYrProMetAlaProProAsnGlnHisValSerProValGln 316
DB 535 -ACTGACCCCATCAG----- 549
QY 316 aLysYrIleLeuLysAspSerPheSerIlePheCysGluThrGlyYrGluLeuGlu 336
DB 550 -----GAGGCCACGAG 560
QY 336 nGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspAr 356
DB 561 GGCGCTCT----- 567
QY 356 gProMetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyAr 376
DB 567 ----- 567
QY 376 vAlaGluYrIleThrGlyProGlyValThrThrYrLysAlaValIleGlnYrSerCys 396
DB 568 ---GAGGCATCAACGACCTGAGACAACTGGCAAG-----GTCCAGAACCACTG 617
QY 396 sGluGluThrPheYrThrMetLysValAsnAspGlyLysYrValCysGlnAlaAspG 416
DB 618 CCAAGAGCCCTATTAT---CAGCCCGCGCAGCAGGAGGACATCACTGTGTCACACCCACG 674
QY 416 yPheTrpThrSerSerLysGlyLysSerLeuProValCysGluProValCysGly 435
DB 675 GACCTGAAAGACAGACAGAGATGGGAGAGGTTCTTCACTGTATGCTGTGTGCGGAGG 734
QY 436 -----LeuSerAlaArgThrThrGlyArgIleYrGlyGlnLys 450
DB 735 GCCAGTACACCCCATTCGCCCAATGACAGAGC-----CTGGGTCTTCCAG 782
QY 450 sAlaLysProGlyAspPheProTrpGlnValLeuIleLeuGlyGlyThrThrAlaAlaG 470
DB 783 AGCCAGAGCTGGGCAACTTCCCTGGCAAGCCTTCAACAGATGACAGCGCGGTGGCGCG 842
QY 470 yAlaLeuLeuYrAspAsnTrpValLeuThrAlaAlaHisAlaValYrGluLys 489
DB 843 GGCCCTGTGGGACAGATGATCTCTCACTGTGCCACACCATATTCACCCCAAGAGAG 902
QY 490 -----HisAspAlaSerAlaLeuAspIleArgMetGly 501
DB 903 TGTTCCTCAGAGAACCAACAGATGATGTCTTGGGCAACAGCCATGATAGA 962
QY 501 rLeuLysArgLeuSerProHisYrThrGlnAlaTrpSerGlyAlaValPheIleHisG 521
DB 963 GATGCTGAACACTGGGAACACACCTGTC-----CACCGTGTGCTTGCACCC 1010
QY 521 uGlyYrThrHisAspAla-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAs 539
DB 1011 CGACTAGCCTCAGATGAGATCCCATTAATCTTAAAGCGGAGATGCGCTCTCGAGCTGCA 1070
QY 539 nAsnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGlnAlaG 559

```

Db 1071 GCACAGATCCCGCTGGCCCAAGCTCCCGGTCTGCTGCC-----GATAATGA 1124
Oy 559 userPhemeArGThrAspAspIleGlyThrAlaSerGlyTrpGlyLeuThrGlnArgG1 579
Db 1125 GACCCCTACACGAGCGCTTGTGGGTACGTCAGTGGGTGGCATG--GAGATGGG 1181
Oy 579 yPheLeuAlaArGAsnLeuMetGlyValAspIleProIleValAspHisGlnIleCysTh 599
Db 1182 CTGGCTAACTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1241
Oy 599 rAlaAlaTrpGlyLysProProTrpProArgGlySerValThrAlaAsnMetLeuCysAl 619
Db 1242 CCGCTGGCTCCAAAGAGACAGACAGACCGAG--GTGTTTCTTGACAAATATGTTCTGTGT 1298
Oy 619 aGlyLeuGlySerGlyGlyLysAspSerGlyAspArgGlyAspSerGlyValAlaLeuValh 639
Db 1299 TGGGATGAGAGCAAGCAAGCAGTGTCTGCCAGGGGACAGTGGCAGCCTCATGTGTGT 1358
Oy 639 eLeuAspSerGlyThrGlyLysArgTrpPheValGlyGlyIleValSerTrpGlySerMetAs 659
Db 1359 ATGGGACATCATGACCCATCAGCTGCGTGGCCACAGGGCATTTGTCTGTGGGC--ATAAG 1415
Oy 659 nCysGlyGlyAlaGlyGlyTrpGlyValIleTrpThrLysValIleAsnTrpIleProTrpI1 679
Db 1416 GTGTGGCGAAGG-----TATGACTTCTACACCAAGTGTCTGACTATGTGACTGTGAT 1469
Oy 679 eGluAsnIleIleSer 684
Db 1470 CAAGGAGTGTATGAAT 1485

RESULT 17
AAS14092/c
ID AAS14092 standard; DNA; 1691 BP.

AAS14092:
XX 18-DEC-2001 (first entry)
XX
XX Human FCTR5b DNA sequence.
DE
XX
XX Human; FCTR5b; myelogenous leukemia; carcinoma; melanoma; glioma; ds;
XX astrocytoma; congenital neonatal autoimmune thrombocytopenia; infection;
XX neurolycogen; congenital neurodegenerative disorders; nerve trauma;
XX familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
XX demyelinating Gardner syndrome; familial myelodysplastic syndrome;
XX mental health condition; immunological disorder; allergy; infertility;
XX bronchial asthma; Avelino type eosinophilia; lung disease; deafness;
XX reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
XX desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
XX gastric disorders; pancreatic disease; Schistosoma mansoni infection;
XX Spino cerebellar ataxia; Plasmodium falciparum parasitemia; diabetes;
XX Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;
XX Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
XX antiatherogenic; antiasthmatic; antiinfectivity; antiinflammatory;
XX antidiabetic; protozoacide; hepatotropic; virucide; ophthalmological;
XX gynaecological; antiinfectivity; immunostimulant; auditory; haemostatic;
XX gene therapy; FCTR5b; human complement C1R component precursor.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 256..1423
XX FT /*tag= a
XX FT /product= "Human FCTR5b"
XX FT 1..255
XX FT 5'UTR
XX FT /*tag= b
XX FT 3'UTR 1424..1691
XX FT /*tag= c
XX PN MO200166747-A2.
XX PD 13-SEP-2001.

```

```

XX 05-MAR-2001; 2001WO-US07160.
PF 03-MAR-2000; 2000US-186592P.
XX 03-MAR-2000; 2000US-186718P.
PR 06-MAR-2000; 2000US-187293P.
PR 06-MAR-2000; 2000US-187294P.
PR 17-MAR-2000; 2000US-190400P.
PR 07-APR-2000; 2000US-196018P.
PR 03-JAN-2001; 2001US-259548P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Vernet CAM, Fernandes E, Shinkels RA, Herrmann JL, Majumder K;
PI MacDougall J, Mishra V, Mezes PS, Rastelli L,
XX
XX WPI. 2001-596837/67.
DR P-PSDB; AAU08684.
XX
XX Novel polypeptides designated as FCTR5b polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
PT states -
XX
XX Claim 9; Page 94; 215pp; English.
XX
XX The invention relates to human FCTR5b polypeptides, FCTR1-FCTR7, and the
XX nucleic acids encoding them. These sequences are useful for the treatment
XX or prevention of numerous disorders including myelogenous leukemia,
XX carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
XX autoimmune thrombocytopenia, neurological disorders, neurodegenerative
XX disorders, nerve trauma, familial myelodysplastic syndrome,
XX Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
XX myelodysplastic syndrome, mental health conditions, immunological
XX disorders, allergy and infection, bronchial asthma, Avelino type
XX eosinophilia, lung diseases, reproductive disorders, infertility, male
XX and female reproductive disorders, deafness, glycoprotein Ia deficiency,
XX desmoid disease, turcot syndrome, liver cirrhosis, hepatitis C, gastric
XX disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
XX infection Spino cerebellar ataxia, Plasmodium falciparum parasitemia,
XX Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
XX Reis-Bucklers corneal dystrophy. This sequence represents DNA encoding
XX FCTR5b, a protein bearing sequence homology to human complement C1R
XX component precursor.
XX Note: The specification provides the start and stop codons of the coding
XX sequence but the DNA does not appear to encode the stated protein.
XX
XX SQ Sequence 1691 BP; 349 A; 502 C; 484 G; 356 T; 0 other:

Alignment Scores:
Pred. No.: 4,84e-31 Length: 1691
Score: 496.00 Matches: 157
Percent Similarity: 33.67% Conservative: 74
Best Local Similarity: 22.89% Mismatches: 194
Query Match: 13.20% Indels: 261
DB: Gaps: 20

US-09-874-198-2 (1-686) x AAS14092 (1-1691)
Oy 17 ProLeuGlyProLys-TTrpGluProValPheGlyAlaGlyLeuAlaSerProGlyPhePr 36
Db 1509 CCTTGGGCCCAAGAGCTACCCAG-----CAGCTGACATCCCGCGGATACC 1462
Oy 36 oGlyGlyTrpAlaAsnAspGlyGlyLysArgTrpTrpThrAlaProProGlyLysArg 56
Db 1461 AGAGCCGTATGGCAAGGCCAAGAGAGCAGCAGCAGCATCAAGGCTCCAGAGGCTTGC 1402
Oy 56 gLeuArgLeuTrpPheThrHisPheAspLeuGlyLeuSerHisLeuCysGlyTrpAspPh 76
Db 1401 TGTGAGGCTCGTCTTCCAGAGACTTGCAGTGGAGCGTCCAGAGACTGAGGGAGCTC 1342
Oy 76 eValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGlySerThrAs 96
Db 1341 TGTCAATCATCTATTCGTGGTTCGATCCAGCAGGATTCGTGTGTCAGCAGGCTCCC 1282

```

```

QY 96 pthrgluargAlaProGlyLysAspThrPheTySerLeuGlySerSerLeuAspIleTh 116
    ||| |||||::: ||| ||| |||||::: |||
Db 1281 TCTGGCAGACCCCTGTCAGAGAGATTGTATCTCAGAGAGATTTCGGCTGAC 1222
QY 116 rPheArgSerAspTySerAsnGluLysProPheThrGlyPheGluAlaPheTyArgAla 136
    |||||::: ||| ||| |||||::: |||
Db 1221 CTTCGCCACAGACCTTCTCTGGAG----- 1197
QY 136 agluAspIleAspGluCysGlnValAlaProGlyGluAlaProThrCysAspHisScy 156
    |||
Db 1197 ----- 1197
QY 156 shIsAsnHisLeuGlyGlyPheTyCysSerCysArgAlaGlyTyValLeuHisArgAs 176
    ||
Db 1196 -----AA 1195
QY 176 nlysArgThrCysSerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluE 196
    |||
    |||
    |||
Db 1194 CAAG----- 1191
QY 196 userSerProGluTyProArgProTyProLysLeuSerSerCysThrTySerIleSe 216
    |||
    |||
    |||
Db 1190 -----ACTGCCCA 1183
QY 216 rleuGluGluGlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrH 236
    ||| |||||::: |||
Db 1182 CCTCCACAGAGGGCTTCTGGCCCTTAC----- 1155
QY 236 sProGluThrLeuCysProTyAspPheLeuLysIleGlnThrAspArgGluGluHisGl 256
    |||
    |||
    |||
Db 1155 ----- 1155
QY 256 yProPheCysGlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrI 276
    ||| |||||::: |||
Db 1154 -----CAAACCGTGCTGT 1141
QY 276 ethrPheValThrAspGluSerGlyAspHisThrGlyTyTrpLysIleHisTyThrSerTh 296
    ||| |||
    |||
    |||
Db 1140 GAACAT----- 1134
QY 296 rAlaGlnProCysProTyProMetAlaProProAsnGlyHisValSerProValGlnAl 316
    ||| |||||::: |||
Db 1133 -ACTGCCCATCAG----- 1119
QY 316 alysTyriLeuLysAspSerPheSerIlePheCysGluThrGlyTyTrpGluLeuGlu 336
    ||| |||
    |||
    |||
Db 1118 -----GAGGCCAGCAG 1108
QY 336 nGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspAr 356
    |||
    |||
    |||
Db 1107 GGCTCT----- 1101
QY 356 gProMetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyAr 376
    |||
    |||
    |||
Db 1101 ----- 1101
QY 376 gValGluTyriLeThrGlyProGlyValThrThrTyriLysAlaValIleGluTySerCy 396
    ||| ||| |||||::: |||
Db 1100 -----GAGGCCATCAGCAGCTGGAGACAACCTGCCAG-----GTCCAGAACACTG 1051
QY 396 sGluGluThrPheTyThrMetLysValAsnAspGlyLysTyValCysGluAlaAspGl 416
    ||| ||| |||||::: |||
    ||| ||| |||||::: |||
    ||| ||| |||||::: |||
Db 1050 CCAGGAGCCCTATTAT-----CAGGCCCGCAGCAGGCGACCTCAGCTGCACACCCAG 994
QY 416 yPheThrPheThrSerSerLysGlyLysSerLeuProValCysGluProValCysGly-- 435
    ||| ||| |||||::: |||
    ||| ||| |||||::: |||
    ||| ||| |||||::: |||
Db 993 GACCTGGAAAGACAGACAGATGGAGAGGTTCTTCAGTGTATCCCTGTCGCGAGC 934
QY 436 -----LeuSerAlaArgThrThrGlyGlyArgIleTyrcGlyGlyGlnLys 450
    ||| ||| |||||::: |||
    ||| ||| |||||::: |||
    ||| ||| |||||::: |||
Db 933 GCCAGTCACCCCATTTGCCAGATATCAGACGAC-----CTGCTTCTTCCAG 886

```

```

QY 450 salAlaLysProGlyAspPheProTrpGlnValLeuIleLeuGlyGlyThrThrAlaIleGl 470
    ||||| |||||::: ||||| ||||| |||||
    ||||| |||||::: ||||| ||||| |||||
    ||||| |||||::: ||||| ||||| |||||
Db 885 AGCCAGACCTGGGCAACTCCCTGCGCAAGCCTTCACACAGATATCAGGCCGCTGGGCGG 826
QY 470 yAlaLeuLeuTyAspAsnTrpValLeuThrAlaAlaHisAlaValTyrcGluLys-- 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 825 GGCCCTGCTGGGACAGATGATCTCACTGCTGCCACACCATATCAGCCCAAGACAG 766
QY 490 -----HisAspAlaSerAlaLeuAspIleArgMetGly-----Th 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 765 TGTTCCTTCAGAGAACACAGAGTGTGAATGTCTTCTTGCGCCACAGACCATATGCA 706
QY 501 rleuLysArgLeuSerProHisTyThrGlnAlaTrpSerGlyAlaValPheIleHisGl 521
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 705 GATGCTGAACCTGGGACACCATCCTGTC-----CACCGTGTCTGTGACACC 658
QY 521 uGlyTyThrHisAspAla-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAs 539
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 657 CGACTACCGTCAGAAATGAGTCCCATTAACCTTACGGGGACATCGCCCTCGGAGCTGCA 598
QY 539 nAsnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGl 559
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 597 GCACAGCATCCCTGGCCCAACGTCCTCCGGTGTGTGCC-----GATATGA 544
QY 559 userPheMetArgThrAspAspIleGlyThrAlaSerGlyTyTrpGlyLeuThrGlnArgGl 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 GACCTCTACCGCAGGCGCTTGGGCTACGTACGTACGTGGTTGGCATG--GAGATGG 487
QY 579 yPheLeuAlaArgAsnLeuMetTyValAspIleProIleValAspHisGlnLysCysTh 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 486 CTGCTAACTACTGAGCTGAACTGAGCTGAGCTGCTGTAGCTCCACAGGAGCCTCGAA 427
QY 599 rAlaAlaTyrcGluLysProProTyProArgGlySerValThrAlaAsnMetLeuCysAl 619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 426 CGCCTGCTCCAAAGAGACAGACAGACCCAG--GCTTTTCTGACAAATATGTCTGTGT 370
QY 619 aglyLeuGluSerGlyGlyLysAspSerCysArgGlyAspSerGlyValAlaLeuValPh 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 TGGGATGAGACGCCAAAGGACAGTGTCTGCCAGGGGACAGTGCACCTATATGTGCT 310
QY 639 eleuAspSerGluThrGluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAs 659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 309 ATGGACAAATCATGCCCATCATCTGCGTGGCCACGGGATGTGTCTCGGGGCG--ATAGG 253
QY 659 nCysGlyGluAlaGlyGlyLysAspSerCysArgGlyAspSerGlyValAlaLeuValPh 679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 GTGTGGCGAAGG-----TATGACTTTTACACCAAGTGTCTGAGCTATGTGACTGAT 199
QY 679 eGluAsnIleIleSer 684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 CAAGGAGTGTATGAT 183

```

```

RESULT 18
AAx87815
ID AAx87815 standard; cDNA; 3147 BP.
XX
AC AAx87815;
XX
DT 09-NOV-1999 (first entry)
XX
DE Tumour antigen derived gene-15 (TADG-15) cDNA.
XX
KW Tumour antigen derived gene-15; TADG-15; serine protease; human;
KW breast cancer; ovary cancer; carcinoma; diagnosis; ss.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX FT 23..2590
XX FT /tag= a
XX PN
XX
XX MO9942120-A1.

```

PD 26-AUG-1999 .
XX
XX 18-FEB-1999; 99WO-US03436.
XX
XX 20-FEB-1998; 98US-0027337.
XX
XX (UVR-) UNIV ARKANSAS.
XX
XX O'Brien TJ, Tanimoto H;
DR WPI; 1999-527418/44.
DR P-PSDB: AAY06671.
XX
XX A new extracellular serine protease for diagnosis of neoplastic
PT disease
PS Claim 2: Fig 9; 71pp: English.
XX
XX
XX This is the nucleotide sequence of an isolated cDNA that codes for
CC an extracellular serine protease, termed tumour antigen derived
CC gene-15 protein (see AAY06671), that is overexpressed in breast and
CC ovarian carcinomas. The TADG-15 gene can be used as a diagnostic
CC and therapeutic target in ovarian carcinoma and other carcinomas
CC including breast, prostate, lung and colon. The TADG-15 cDNA was
CC isolated from ovarian carcinoma by PCR using primers directed to
CC conserved areas of the serine protease family. The invention
CC also provides: a vector that is capable of expressing DNA encoding
CC TADG-15 protein; host cells selected from bacterial cells
CC (especially Escherichia coli), mammalian cells, plant cells and
CC insect cells; and a method of detecting expression of TADG-15
CC protein using a hybridisation probe.
XX
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Alignment Scores:
Pred. No.: 3,35e-30 Length: 3147
Score: 490.00 Matches: 186
Percent Similarity: 37.81% Conservative: 90
Best Local Similarity: 25.48% Mismatches: 270
Query Match: 13,04% Indels: 184
DB: 20 Gaps: 31

US-09-874-198-2 (1-686) x AAX87815 (1-3147)

QY 29 ATGLeuAlaSerProGlyPhePro---GlyGluTyrAlaAsnSprIngluArgTyr 47
DB 701 CGCTTACCAAGCGCGCGCTTCTGACAGCCCTACCCCGCTGACCGCGTGG 760

48 ThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeu 67
DB 761 GCCCTGCGGGGAGCCGACCTAGTGCATGCTGACCTCAGCTTGGCAGCTTGGC 820

68 LeuSerHisLeuGlyGluTyrAspPheValLysLeuSerSerGlyAlaLys----- 84
DB 821 TCCTGCGAGAGCGGCGGACGCGTGGTACAGCTACACACCTGAGCCCATGAG 880

85 -----ValLeuAlaThrLeuGlyGlnGlnuSerThrAspThrGlnuAlaProGly 102
DB 881 CCCACGCGCTGGCGAGTGTGTGGC-----ACCTAACCTCCCTCTAC 925

103 LysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrSer 122
DB 926 AACCTACCTTCACCTCCGAGACGTCCTGCTCATGCTACATGATACCACTGAG 985

123 AsnGlnLysProPheThrGlyPheGlnuAla---PheTyrAlaAlaGlnuSprIleAspGlu 141
DB 986 CGCGCGCATGCC-----GGCTTGGAGCGACCTCTTCCAGCTGCTGAGATGACAGC 1039

142 CysGlnValAlaAlaProGlyGlnuAlaProThrCysAspHisHisCysHisAsnHisLeuGly 161
DB 1040 TGT----- 1042

QY 162 GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSer 181

DB 1042 ----- 1042
QY 182 AlaLeuCysSerGlyGlnValPheThrGlnuArgSerGlyLeuLeuSerProGlyTyr 201
DB 1043 -----GGAGCGCGCTTACGTAAAGCCACAGGAGCATTCACACCCCTACTAC 1090

202 ProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGlnuGlyPhe 221
DB 1091 CCAGGCCACTACCCACCAACATTCAGTACAGTGGACATTCAGAGCTGACCAACACAG 1150

222 SerValIleLeuAspPheValGlnuSerPheAspValGluThrHis---ProGluThrLeu 240
DB 1151 CATGTAGAGTGAAGCTTCAATTCCTTACACCTGACCGTGGACCGCGCTGGGCGAC 1210

241 CysProTyrAspPheLeuLysIleGlnThrAspArgGlnuGlnuHisGlyProPheCysGly 260
DB 1211 TGCCCCAGGAGTACCTGGAGATCAATGGGGAGAA-----TACTCCGA 1255

261 LysThrLeuProHisArgIleGluThrLysSerAspThrValThrIleThrPheValThr 280
DB 1256 GAGAGTCCAGTTCCTGCTGACACCAACAGACACAGATGACAGTTCCTGCTGACCTA 1315

281 AspGluSerGlyAspHisThrGlyTyrLysIleHisTyrThrSer-----ThrAlaGln 298
DB 1316 GATCAGTCTTACACCGACACCGCGCTTACGTGATGATCTGCTTACGACACTGAC 1375

299 ProCysProTyrProMetAlaProProAsnGlnHisValSerProValIleAlaLysTyr 318
DB 1376 CCATGGCCG-----GGGCACTGACGCGCGGCGGCTGT 1414

319 IleLeuLys-----AspSerPhe 324
DB 1415 ATCCGAGAGAGCTGCGCTGATGCTGCGGCGGACCTGACCGACCGACCGATGAGCTC 1474

325 SerIlePheCysGlnuThrGlyTyrGlnuLeuGlnuGlnuHisLeuProLeuSerPhe 344
DB 1475 AACGTGAGTTGGACCGCGGACCGACGTTACGTGACACACAGTGTGCAAGCCCTC 1534

345 ThrAlaValCysGlnuLys-----AspGlySerThrAspArgProMetPro 359
DB 1535 TTCTGGCTCTGCGACAGTGTGAACGACTGCGGAGACMACAGCGAGCAGCAG----- 1585

360 AlaCysSerIleVal----- 364
DB 1586 GGGTGCAGTGTGTCGCGCCACAGCTTCAAGTGTCCAAAGGAAGTGCCTCGAAGAC 1645

365 -----AspCysGlyPro---ProAspAspLeuProSerGlyArg 376
DB 1646 CAGCAGTGCAGATGCGAAGACAGCTGTGGGAGCGGTGCGAGAGAGCGCTCGCCCAAG 1705

377 ValGlnuTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnuTyrSerCys 396
DB 1706 GTGACAGTGCATGCTGTGCAACACACCC-----TACCGCTGC 1744

397 GlnuGluThrPheTyrThrMetLysValAsn-----AspGlyLysTyrValGlyGlu 413
DB 1745 CTCATGGGCTGCTGTTGAGCAAGGCGACCTGAGTGTGAGGAGAGAGACTGT--- 1801

414 AlaAspGlyPheThrPheSerSerLysGlnuLysSerLeuProValCysGlnuProVal 433
DB 1802 AGCGAGCGC-----TCAGATGAGAAAGAC-----TGCAGC----- 1831

434 CysGlyLeuSerAlaArgThrThrGlyGlnuArgIleTyrGlyGlnuLysAlaLysPro 453
DB 1832 TGGGCGCTCGGCAATTCACAGAGAGCGCTGTTGTGGGCGCGAGATGCGAGTTGAG 1891

454 GlyAspPheProThrGlnuValLeuIle-----LeuGly---GlyThrThrAlaGly 470
DB 1892 GCGAGTGGCGCTGCGAGGTAAAGCTGATGCTGTGGGCGAAGGCGCAATCTGGCGTCT 1951

471 AlaLeuLeuTyrAspAsnThrValLeuThrAlaAlaHisAlaValTyrGlnuGlnuLys--- 489

[illegible]

```

PR      03-FEB-2000; 2000US-017998Z.
PR      18-FEB-2000; 2000US-018354Z.
PR      22-JUN-2000; 2000US-0213124.
PR      26-JUL-2000; 2000US-0220970.
PR      08-SEP-2000; 2000US-0657986.
PR      22-SEP-2000; 2000US-0234840.
XX
PA      (CORV-) CORVAS INT INC.
XX
PI      Madison EL, Ong EO, Yeh J:
XX
DR      WPI: 2001-488877/53.
XX      P-PDSB: AAEO6930.
PT
PT      Novel single chain polypeptide comprising protease domain of type-II
PT      membrane-type serine protease or its catalytically active portion
PT      useful for treating and preventing cancer and tumor
XX
PS      Claim 14; Page 191-195; 256pp; English.
CC
CC      The invention relates to transmembrane serine proteases and their
CC      corresponding nucleotides and the protease domain of a type-II
CC      membrane-type serine protease (MTSP). MTSP is useful for identifying
CC      compounds that modulate or inhibits its proteolytic activity and for
CC      formulating a medication for treating neoplastic disease. MTSP and
CC      its corresponding nucleotides are useful in preventing or treating
CC      tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC      ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC      is useful as a diagnostic marker for tumour development, growth and/or
CC      progression and as immunogens to generate antibodies that specifically
CC      bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC      gene therapy. The present sequence is a DNA encoding human MTSP1
CC      protein (also called matrilysin).
XX
SQ      Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other:
XX
Alignment Scores:
Pred. No.:          3..j5e-30          Length:        3147
Score:              490.00             Matches:         186
Percent Similarity: 37.81%             Conservative:    90
Best Local Similarity: 25.48%           Mismatches:     270
Query Match:        13.04%             Indels:         184
DB:                 22                  Gaps:           31
US-09-874-198-2 (1-686) x AADJ3113 (1-3147)
OY      29 ArgLeuAlaSerProGlyPhePro---GlyGLuTYrAlaAsnAspGlnLarGArGrTP 47
       |||
Db      701 CGCTTCACGCCAGCGGGCTTCCCTGCAGCACCCCCCTCATCCCGCGTGCAAGG 760
OY      48 ThrLeuThrAlaProProGlyTYrATrGLeuArgLeuTYrPheThrHisPheAspLeuGlu 67
       |||
Db      761 GCCCTGGGGGGGAGCAGCGGCATCAGTGTAGGCTACCTCCGCAGCTTTACCTTGCG 820
OY      68 LeuSerHisLeuCysGLuTYrAspPheValLysLeuSerSereLYALaLys-----84
       |||
Db      821 TCCTGCAGCAGAGCGCGGCAGCACCTGTGCAGCTACTACAACACCCTGAGCCCAGTAGAG 880
OY      85 -----ValLeuAlaThrLeuCysGLuGlnGluSerThrAspTHrGlnLarGlaProGly 102
       |||
Db      881 CCCACGCCCTGTGGCAGTGTGTGGC-----ACCTAACCTCCCTTCATCC 925
OY      103 LysAspThrPheTyTrSerLeuGlnGlySerSerLeuAspLleThrPheaGyrSerAspTYrSer 122
       |||||
Db      926 AACCTGACCTTCACATCTCCGACGAAGCTGTGTGATACACTAATAACAAACATGAG 985
OY      123 AsnLyuSprProPherThrGlyPheGluAla---PheTYrAlaAlaGlaAspLlleAspGlu 141
       |||
Db      986 CGGCGGCGATCCC-----GGCTTTGAGGCGCACCTTTCCAGCTGCTAGGATGAGCAGC 1038
OY      142 CysGlnAlaLaProGlyGluAlaProThrCyAspRHisHisCysHisAsnHisLeuGly 161
       |||

```

```

Db 1040 TGT----- 1042
QY 162 GLYPheTYrCYsserCYsArGAlaGlyTYrValLeuHisArGsnLYsArGThrCYsSer 181
Db 1042 ----- 1042
QY 182 AlaLeuCYsSerGlyGlnValPheThrGlnArGSerGlyGluLeuSerSerProGluTYr 201
1043 -----GGAGCGCGCTTACGTAAGCCAGGGGACATTCAACAGCCCTTACTAC 1090
QY 202 ProArGProTYrProLYsLeuSerSerCYsThrTYrSerLleSerLleuGluGluGlyPhe 221
1091 CCAGGCCACTACCCACCACCAATTCATGACATGGAACCATTTAGAGGCCCAACAACACAG 1150
QY 222 SerValLleLeuAspPheValGluSerPheAspValGluThrHis---ProGluThrLeu 240
1151 CATGTAGAGTGAGCTTCAAAATCTTCTACCTCTGAGGCCCGCGCTGCGGGGCACC 1210
QY 241 CysProTYrAspPheLeuLYsLleGlnThrAspArGGLuGluHisGlyProPheCYsGly 260
211 TCCCCAGAGACTACGTGAGATCAATGGGAGAAA-----TACTCGCGA 1255
261 LysThrLeuProHisArGlyLleGluThrLysSerAsnThrValThrLleThrPheValThr 280
1256 GAGAGTCCAGTTCGTGTCACACCAACACAGACAAAGATCAAGTTCAGTTCACACTCA 1315
QY 281 AspGluSerGlyAspHisThrGlyTYrPLysLleHisTYrThrSer-----ThrAlaGln 298
1316 GATCACTCCACACCAACACCGGCTTCTTACGTAATCACTCTCTCAACACTCCAGTGAC 1375
QY 299 ProCYsProTYrProMetaLarProProAsnGlyHisValSerProValGlnAlaLYsTYr 318
1376 CCATGCCCG-----GGCAGCTTCACGTGCCGACGGGGCGGTGT 1414
QY 319 IleLeuLYs-----AspSerPhe 324
1415 ATCCGGAAGAGAGCTGCGCTGTATGAGCTGGGCCACTGCACCGACACAGCGATGAGCTTC 1474
QY 325 SerLlePheCYsGluThrGlyTYrGluLeuGlnGlyHisLleuProLeuLYsSerPhe 344
1475 AACTGCAGTTGGGACCGCGGCCACAGTTCACTGCAGAACAAAGTTCTGCAGAGCCCTC 1534
QY 345 ThrAlaValCYsGlnLYs-----AspGlySerThrAspArGProMetaPro 359
1535 TTCTGGGTCTGCACAGATGTGAACGACTGGGAGACACAGCAGCAGCAGCAG----- 1585
QY 360 AlaCYsSerLleVal----- 364
586 GGGTGCAGTGTCCGGCCACAGACTTCAGGTTCCTCAATGGGAAGTGCCTTCGAANAAGC 1645
365 -----AspCYsGlyPro---ProAspAspLeuProSerGlyArg 376
1646 CACGAGTGCATGGGAAGGAGCAGCTGTGGGAGGCGTCCAGCAGAGCCCTCTCCCAAG 1705
QY 377 ValGluTYrLleThrGlyProGlyValThrTYrLYsAlaValLleGlnTYrSerCYs 396
1706 GTGAAGCTGCTCACTGTACCAAAACACACC-----TACCGCTGC 1744
QY 397 GluGluThrPheTYrThrMetLYsValAsn-----AspLYsTYrValLYsGlu 413
1745 CTCGAATGGGCTGCTTGAAGCAAGGCCAACCCTGAGTGTACGGGAAGGAGGACTGT--- 1801
QY 414 AlaAspGlyPheThrPheSerSerLYsGlyLysSerLeuProValCYsGluProVal 433
1802 AGCGAGCGC-----TCAGATGAGAAGGAC-----TCCGAC----- 1831
QY 434 CysGlyLeuSerAlaArGThrThrGlyGlyArgLleTYrGlyGlyGlnLYsAlaLYsPro 453
1832 TCTGGGCTGGGTCATTCACAGACAGAGCTCGTGTGTGGGGCAGCGATGGCATGAG 1891
QY 454 GlyAspPheProTYrPheValLeuLle-----LeuGly---GlyThrThrAlaAlaGly 470
1892 GCGGAGTGGCCCTGGAGGTAAGCTGCATGCTGTGGCGCAGGGCCACATCTCGGTGCT 1951

```

```

QY 471 AlaLeuLeuTYrAspAsnTrpValLeuThrAlaAlaHisAlaValTYrGluGlnLYs--- 489
1952 TCCCTATCTCTCCCAACATGGCGTGTCTGGCGGCACATGCTACTACATGACAGAGA 2011
QY 490 -----HisAsp----- 491
2012 TTCAGTACTACAGACCCACGAGTGAGCGGCTTCTGGGCTGACACCAAGACCCAG 2071
QY 492 AlaSerAlaLeuAspLleArGmetGlyThrLleLYsArGLeuSerProHisTYrThrGln 511
2072 CGCAGCGCCCTGGGTGAGGAGCGCCAGGCTCAAGCGCATATCTCCAC----- 2122
QY 512 AlaTrpSerGluAlaValPheLleHisGluGlyTYrThrHisAspAlaGlyPheAspAsn 531
2123 -----CCCTTTCATAGACTTCACTTCGACAT 2152
QY 532 AspLleAlaLeuLleLYsLeuAsnLYsValValLleAsnSerAsnLleThrProLle 551
2153 GACATCGCGCTGTGAGGTGAGAAACCGGACAGATACACCTCCATGTGGCGCCATC 2212
QY 552 CysLeuProArGlySGluAlaGluSerPheMetArGThrAspAspLleGlyThrAlaSer 571
2213 TGCCTCCG-----GAGCCTCCCATGTCTTCCTCCCGGCAAGGCCATCTGGGTCACG 2266
QY 572 GlyTrpGlyLeuThrGlnArGlyPheLeu---AlaArGAsnLeuMetLYrValAspLle 590
2267 GCGTGGGACACACCCAGATGAGGAGGACTGGCGGCTATCTCGCAAAAGGTGAGATC 2326
QY 591 ProLleValAspHisGlnLYsCYsThrAlaAlaTYrGluLYsProProTYrProArGly 610
2327 CGGCTCATACACGACACCTGCGAAGCCCTCTCCGACG----- 2368
QY 611 SerValThrAlaAspMetLeuCYsAlaGlyLeuGluSerGlyLYsAspSerCYsArg 630
2369 CAGATACGCGCGCATGTGCTGTGCTTCACGCGGCGGTGAGCTGTCCAG 2428
QY 631 GlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGluArGTrpPheValGly 650
2429 GGTGATTCGGGGGAGCCCTGTTCACGCGTGAAGGCGGAT---GGGCGATCTTCCAGGCGC 2485
QY 651 GlyLleValSerTrpGlySerMetAsnCYsGlyGluAlaGlyGlnTYrGlyValTYrThr 670
2486 GGTGTGTGAGTGGGGAAC---GGCTCGGCTCACAGAGAACAGCAGCGCTGTACACA 2542
QY 671 LysValLleAsnTYrLleProTrpLleGlu 680
2543 AGGCTCCTCTGTTCGGGACTGATCAAA 2572
Db 2543 AGGCTCCTCTGTTCGGGACTGATCAAA 2572

RESULT 20
AADI3155
ID AADI3155 standard; DNA; 3147 BP.
XX
AC AADI3155;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human membrane-type serine protease (MTSP) 1 protease domain DNA.
XX
KW Human; transmembrane serine protease; Membrane-type serine protease;
KW MTSP; protease domain; neoplastic disease; tumor; cancer; cytosolic;
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW matiprase; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1865..2590
FT /tag= a
FT /product= "Human transmembrane serine protease
FT (MTSP) 1 protease domain"
FT /note= "CDS does not include start codon"
FT

```

XX WO200157194-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 02-FEB-2001: 2001WO-US03471.
 XX
 XX 03-FEB-2000: 2000US-0179982.
 XX 18-FEB-2000: 2000US-0183542.
 XX 22-JUN-2000: 2000US-0213124.
 XX 26-JUL-2000: 2000US-0220970.
 XX 08-SEP-2000: 2000US-0657986.
 XX 22-SEP-2000: 2000US-0234840.
 XX
 XX (CORV-) CORVAS INT INC.
 XX
 XX Madison EL, Ong EO, Yeh J;
 XX
 XX WPI: 2001-488877/53.
 XX P-PSDB: AAE06936.

PT Novel single chain polypeptide comprising protease domain of type-II
 PT membrane-type serine protease or its catalytically active portion
 PT useful for treating and preventing cancer and tumor

Example 6; Page 225-227; 256pp; English.

CC The invention relates to transmembrane serine proteases and their
 CC corresponding nucleotides and the protease domain of a type-II
 CC membrane-type serine protease (MSP). MSP is useful for identifying
 CC compounds that modulate or inhibit its proteolytic activity and for
 CC formulating a medicament for treating neoplastic disease. MSP and
 CC its corresponding nucleotides are useful in preventing or treating
 CC tumors or cancers such as lung carcinoma, colon adenocarcinoma and
 CC ovarian carcinoma. In diagnostics and in hybridisation assays. MSP
 CC is useful as a diagnostic marker for tumor development, growth and/or
 CC progression and as immunogens to generate antibodies that specifically
 CC bind to it. MSP DNA is useful in a yeast two-hybrid system and in
 CC gene therapy. The present sequence is a DNA encoding protease domain of
 CC human MSP1 protein (also called matrilysin).

XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Alignment Scores:

Pred. No.: 3,35e-30 Length: 3147
 Score: 490.00 Matches: 186
 Percent Similarity: 37.81% Conservative: 90
 Local Similarity: 25.48% Mismatches: 270
 Match: 13.04% Indels: 184
 Gaps: 31

US-09-874-198-2 (1-686) x AAD13155 (1-3147)

QY 29 ArgLeuAlaSerProGlyPhePro---GlyGluTyrAlaAsnAspGlnGluArgTyr 47
 DB 701 CGCTGACACCGCCGCGTCCCTGACAGCCCTACCGCTCATGCCGCTCCAGTGG 760
 QY 48 ThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuGlu 67
 DB 761 GCCCTGCGGGGAGCCGACCTCAGTGTGAGCTCACTCCGACGCTTGGACCTTGGC 820
 QY 68 LeuSerHisLeuGluSerGluTyrAspPheValIysLeuSerSerGlyAlaLys----- 84
 DB 821 TCCTGGAGAGCGGGGAGCGGACGCTGAGCGTGTACACACCCCTGAGCCCATGAG 880
 QY 85 -----ValLeuAlaThrLeuGluGlnGluSerThrAspThrGluArgAlaProGly 102
 DB 881 CCCACAGCCCTGCGTGCAGTTGTGTGGC-----ACCTAACCTCCCTCCTAC 925
 QY 103 LysAspThrPheTyrSerLeuGluSerSerLeuAspIleThrPheArgSerAspTyrSer 122
 DB 926 AACCTGACCTTCCATCTCCGAGAACGTCGTCTCATCATCACTGATTAACCAACTGAG 985

QY 123 AsnGluLysProPheThrGlyPheGluAla---PheTyrAlaAlaGluAspIleAspGlu 141
 DB 986 CGCGGCGCATCC-----GGCTTGAAGCCACCTTCTCCAGCTGGCTAGAGACAGAC 1039
 QY 142 CysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGly 161
 DB 1040 TGT----- 1042
 QY 162 GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSer 181
 DB 1042 ----- 1042
 QY 182 AlaLeuGluSerGlyGlnValPheThrGlnArgSerGlyLeuLeuSerSerProGluTyr 201
 DB 1043 -----GGAGCGCGCTTACGTAAGCCAGGAGGACATTCACAGCCCTTCTAC 1090
 QY 202 ProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGluGlyPhe 221
 DB 1091 CGAGGCGACATCCACCCCAACATTGACTGTCACATGACATGACATGACGCGCCACACAC 1150
 QY 222 SerValIleLeuAspPheValGluSerPheAspValGluThrHis---ProGluThrLeu 240
 DB 1151 CATGTAGAGGTGAGCTCAAAATTCTTCTACCTGTCGAGCCGCGTGCCTGGCGCAC 1210
 QY 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnIleGlyProPheCysGly 260
 DB 1211 TGCCCAAGGACATCACTGAGATCAATGGGAGAAA-----TACTGCGGA 1255
 QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr 280
 DB 1256 GAGAGGTCACGATCTGTCGACACACGACACACAAAGATCAAGTACAGTCCCTCCACTGA 1315
 QY 281 AspGluSerGlyAspHisThrGlyTyrLysIleHisTyrThrSer-----ThrAlaGln 298
 DB 1316 GATCAGTCTTACACGACGACGCGGCTTCTTACGTAATACCTCTCCACGACTCCACTGAC 1375
 QY 299 ProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyr 318
 DB 1376 CATTGCCG-----GGCAGTTCACTGCCGACGGGCGGTGT 1414
 QY 319 IleLeuLys-----AspSerPhe 324
 DB 1415 ATCCGGAAGAGAGCTGCGCTGTGATGCTGGCGCCGACTGCACGACACAGCATGAGCTC 1474
 QY 325 SerIlePheGluGluThrGlyTyrGluLeuLeuGlnGlyHisLeuProLeuLysSerPhe 344
 DB 1475 AACTGCAAGTTGCGACCGCGCCGACAGTTCACTGCAAGAAAGTTTTCGCAAGCCCTTC 1534
 QY 345 ThrAlaValCysGlnLys-----AspGlySerTyrPaspArgPrometPro 359
 DB 1535 TTCTGGGTCTGCGACAGTGTGAACGACTGCGGAGACACACGACGACGACGACGACG 1585
 QY 360 AlaCysSerIleVal----- 364
 DB 1586 GGGTGCAGTGTGTCGCGCCGACGCTTCAAGTGTTCGAATGGAAGTGCCTTCGAAAGC 1645
 QY 365 -----AspCysGlyPro---ProAspAspLeuProSerGlyArg 376
 DB 1646 CAGCACTGCANAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1705
 QY 377 ValGluTyrIleThrGlyProGlyValIleThrTyrIleValAlaIleGlnTyrSerCys 396
 DB 1706 GTTAAGCTGCTCACTGTATCAAAACACC-----TACCGCTGC 1744
 QY 397 GluGluThrPheTyrThrMetLysValAsn-----AspGlyLysTyrValCysGlu 413
 DB 1745 CTCGAATGGGCTGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1801
 QY 414 AlaAspGlyPheThrPheThrSerSerLysGlyGluLysSerLeuProValCysGluProVal 433
 DB 1802 AGCGAGCGC-----TCAGATGAGAGGAGC-----TCCGAC----- 1831
 QY 434 CysGlyLeuSerAlaArgThrThrGlyGlyArgIleTyrGlyGlyGlnLysAlaLysPro 453

XX WPI: 2001-381031/40.

XX Novel extracellular serine protease, termed tumor antigen-derived gene
PT 15 protein overexpressed in carcinomas and DNA encoding it, for
PT diagnosis, treatment, prevention of cancer, particularly breast,
PT ovarian cancer

XX Example 8; Page 102-103; 130pp; English.

XX The present invention relates to human tumor antigen-derived gene 15
CC (TADG-15) protein and coding sequence (see AAH23601 and AB98500).
CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is
CC over-expressed in ovarian tumors. TADG-15 protein or its fragments of
CC 9-20 residues that lack TADG-15 protease activity are useful for
CC vaccinating an individual against TADG-15, having, suspected of having or
CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a
CC diagnostic or therapeutic target in cancer. TADG-15 was cloned and
CC expressed in the opposite orientation such that an antisense RNA molecule
CC (the present sequence) was produced. The present sequence is useful for
CC hybridizing to the complementary RNA in a cell and thereby inhibiting
CC translation of TADG-15 into protein.

SO Sequence 3147 BP; 583 A; 958 C; 952 G; 654 U; 0 other:

Alignment Scores:

Pred. No.:	3.35e-30	Length:	3147
Score:	490.00	Matches:	186
Percent Similarity:	37.81%	Conservative:	50
Best Local Similarity:	25.48%	Mismatches:	270
Query Match:	13.04%	Indels:	14
DB:	22	Gaps:	31

US-09-874-198-2 (1-686) x AAH23609 (1-3147)

QY 29 ArgLeuAlaSerProGlyPhePro---GlyGluTyrAlaAsnAspGlnGluArgTyr 47
DB 2447 CGCTTCACCGCGCGCGCTTCCTGACAGCCCGCTACCCCGCTGACCGCGCTGCG 2388
QY 48 ThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspGlu 67
DB 2387 GCCCTGCGGGGGAGCCCGCAGCTAGCTGCTAGCTCAGCTTCCGCGAGCTTGGACCTTGGC 2328
QY 68 LeuSerHisLeuGluGluTyrAspPheValLysLeuSerSerGlyAlaLys----- 84
DB 2327 TCCTGCAGACGCGCGACGCGCTGACGCTGACGCTGACACACCCGACGCCCATGAG 2268
QY 85 -----ValLeuAlaThrLeuGluGluGlnGlnGlnSerThrAspThrGluAlaProGly 102
DB 267 CCCCACGCGCTGCTGAGTGTGTGCG-----ACCTAACCTCCCTGCTAC 2223
QY 103 LysAspThrPheTyrSerLeuGluSerSerLeuAspThrPheArgSerAspTyrSer 122
DB 2222 AACCTGACCTTCACCTCCTCCAGAACGCTCTCTCATCAGCTATGATACCAACACTGAG 2163
QY 123 AsnGlnLysProPheThrGluGluAla---PheTyrAlaAlaGluAspLeu 141
DB 2162 CGCGCGCATGCC-----GCGTTGAGGCGACACTTCCAGCTGCTAGATGAGCAGC 2109
QY 142 CysGlnValAlaProGluGluAlaProThrCysAspHisHisCysHisAsnHisLeuGly 161
DB 2108 TGT----- 2106
QY 162 GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgLysAlaGlyThrCysSer 181
DB 2106 ----- 2106
QY 182 AlaLeuGluSerGlyGlnValPheThrGlnArgSerGlyLeuLeuSerProGlyTyr 201
DB 2105 -----GGAGGCGCGCTTACGTAAGAGCCAGGGGACATTCACACCCCTGCTAC 2058
QY 202 ProArgProTyrTyrLysLeuSerSerCysThrTyrSerThrSerLeuGluGluGlyPhe 221

DB 2057 CCAGGCGCACTACCCACCAACATTCAGCTGACATGAGACATTCAGAGTGGCCACCAACAG 1998
QY 222 SerValIleLeuAspPheValGluSerPheAspValGluThrHis---ProGluThrLeu 240
DB 1997 CATGTAAAGTAGAGCTTCAATTCCTTACCTGCTGAGACCGCGGCTGCTGCGGAC 1938
QY 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGly 260
DB 1937 TGCCCAAGAGCTAGCTGAGATCATGAGAGAA-----TACTGCGGA 1893
QY 261 LysThrLeuProHisArgIleGluThrLysSerAspThrValThrThrPheValThr 280
DB 1892 GAGAGGTCCAGTTCGTGTACACCAACCAACAGATGACAGTTCCTTCACCTA 1833
QY 281 AspGluSerGlyAspHisThrGlyThrLysIleHisTyrThrSer-----ThrAlaGln 298
DB 1832 GATCACTCTACACGACACCGCGCTTACGATGATACCTCTCTGACAGCTCAGTGAC 1773
QY 299 ProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyr 318
DB 1772 CCATGCGCG-----GGCAGTTCAGCTGCGCGACGCGGCGGTGT 1734
QY 319 IleLeuLys-----AspSerPhe 324
DB 1733 ATCCGAGAGAGCTCGCTGTGTATGCTGCGCGACCGACCGACCGACGATGAGCTC 1674
QY 325 SerIlePheCysGluThrGlyTyrGluLeuGlnGlnHisLeuProLeuLysSerPhe 344
DB 1673 AACTGCAATTCGACCGCGCGCGACGCTCAGCTGACGACAAACAGTTCGCAAGCCCTC 1614
QY 345 ThrAlaValCysGlnLys-----AspGlySerThrAspArgProMetPro 359
DB 1613 TTCTGGCTCTGCGACAGTGTGAACGACTGCGGAGACACAGCAGCAGCAGCAG 1563
QY 360 AlaCysSerIleVal----- 364
DB 1562 GGGTGCAGTTTCGCGCGCCAGACCTTCAGGTGTCCATGGAAGTCCCTGCAAAAC 1503
QY 365 -----AspCysGlyPro---ProAspAspLeuProSerGlyArg 376
DB 1502 CAGCACTGCATGGAAGAGACGACTGTGGGAGCGGTCCGAGAGGCGCTCGGCCAAG 1443
QY 377 ValGluTyrIleThrGlyProGluValThrThrTyrLysAlaValIleGlnTyrSerCys 396
DB 1442 GTCAAGCTGACTGTTGCAACAAACACC-----TACCCTGCG 1404
QY 397 GluGluThrPheTyrThrIleLysValAsn-----AspGlyLysTyrValCysGlu 413
DB 1403 CTCATAGGCTCTGCTTGAAGCAAGGCAACCTGAGTGTGAGGAGAGACTGT--- 1347
QY 414 AlaAspGlyPheThrPheSerSerLysGlyGluLysSerLeuProValCysGluProVal 433
DB 1346 AGCGAGCGC-----TCAGATGAGAAAGAC-----TGCAGC----- 1317
QY 434 CysGlyLysSerAlaArgThrThrGlyGluArgIleTyrGlyGlyLysAlaLysPro 453
DB 1316 TGGGGCTCGGCGCATTCACGAGACAGGCTCGTGTGTGGGGCGAGGATCGCATGAG 1257
QY 454 GlyAspPheProGlnValLeuIle-----LeuGly---GlyThrThrAlaAlaGly 470
DB 1256 GCGAGTGGCGCTGGAGGTAAAGCTGATGCTGTGGGAGGCGCAACATCTGGCGTGT 1197
QY 471 AlaLeuLeuTyrAspAsnThrValLeuThrAlaAlaHisAlaValTyrGluGlnLys 489
DB 1196 TCCCTATCTCTCCCACTGCGTGTGTGTGCGGCACTGCTACATGATGACAGAGA 1137
QY 490 -----HisAsp----- 491
DB 1136 TTCAGTACTACAGCCCAAGAGTGGAGCGCTTCTGAGTTCGACAGCAGAGACAG 1077
QY 492 AlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGln 511
DB 1076 CCGAGCGCGCTGCGGTGAGAGCGCAGCGCTCAAGCGCATCATCTCCAC----- 1026

```

QY 512 AlATpSerGluAlaValPheIleHisGluGlyTyrThrHisAspAlaGlyPheAspAsn 531
DB 1025 -----CCCTTCTCAATGACTTCACCTTCGACTAT 996
QY 532 AspIleAlaLeuIleLysLeuAsnLysValValIleAsnSerAsnIleThrProIle 551
DB 995 GACATCGCCCTCGTACGTGAGAAACCCGACAGATACAGCTCCATGTCGGCCCATC 936
QY 552 CysLeuProArgLysGluIaGluSerPheMetArgThrAspAspIleGlyThrAlaSer 571
DB 935 TGCCTGCCCC-----GACCCTCCCATCTGCTTCCTCCGCGCAAGCCATCTGGGTACAG 882
QY 572 GATTTGGLyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetLysValAspIle 590
DB 881 GGCTGGGACACACCCAGATATGAGGCACTGGCGCCCTATCTCGAANAAGGTGAGATC 822
QY 591 ProIleValAlaSPHisGlnLysCysThrAlaAlaTyrGlnLysProProTyrProArgGly 610
DB 821 CCGCTCATCAACACACACACCTGCGAGAACCTCTCCGACG-----780
QY 611 SerValThrAlaAsnMetLeuGlyAlaGlyLeuGluSerGlyGlyLysAspSerCysArg 630
DB 779 CAGATCACCCCGCATGATGTCGCTTCCTCAGCGCGCTGACCTCCAG 720
QY 631 GlyAspSerGlyValAlaLeuValPheLeuAspSerGlyThrGlnArgThrPheValGly 650
DB 719 GGTGATTCGGGGGACCCCTGTCACAGCTGAGCGGAT---GGCGGATCTTCCAGGCC 663
QY 651 GlyIleValSerTrpLysSerMetAsnCysGlyGlnAlaGlyGlnTyrGlyValTyrThr 670
DB 662 GGTGTGTGTCAGCTGGGGAAC---GGCTGCGCTCAGACAGAACACGACGCGTGTACACA 606
QY 671 LysValIleAsnTyrLleProTrpIleGlu 680
DB 605 AGGCTCCCTCTGTTTCGGGACTGATCANA 576

RESULT 23
AAA88492
ID AAA88492 standard; cDNA; 2955 BP.
XX
AC AAA88492;
XX
DT 22-JAN-2001 (first entry)
XX
DE Human matrilysin (truncated form) cDNA.
XX
XX Matrilysin: serine protease; human; breast cancer; pre-malignancy;
XX clinical keratosis; leukoplakia; Barrett's epithelium;
XX columnar metaplasia; ulcerative colitis; Bowenoid papulosis;
XX adenomatous colorectal polyp; Oesophag erythroplasia;
XX vulvar intraepithelial neoplasia; tumour; metastasis; therapy; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH primer_bind /*tag= a
XX FT complement(33..54)
XX FT primer_bind 2843..2863
XX FT /*tag= b
XX FT CDS 358..2409
XX FT /*tag= c
XX FT sig_peptide 358..450
XX FT /*tag= d
XX FT mat_peptide /*note= "putative signal peptide"
XX FT 451..2406
XX FT /*tag= e
XX
XX WO200053232-A1.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000MO-US06111.

```

```

XX
XX 12-MAR-1999: 99US-0124006.
XX
XX (GEOU ) UNIV GEORGETOWN.
XX
XX PI Dickson RB, Lin C, Johnson M, Wang S, Emyedy I;
XX WPI: 2000-594268/56.
XX P-PSDB: AAB19551.
XX
XX Treating malignancies, premalignant and pathologic conditions in a
XX subject, comprises administering matrilysin modulating agent
XX
XX Claim 9: Fig 9; 116pp; English.
XX
XX The present sequence is that of cDNA coding for the truncated form
XX (see AAB19551) of human matrilysin, a trypsin-like protease. The
XX cDNA was isolated from T-47D human breast cancer cells by RT-PCR.
XX amplification using primers (see AAA88496-97) based on SCM19 cDNA.
XX cDNA encoding the full-length form of matrilysin is given in AAA88493.
XX
XX Either form can be incorporated into a vector and used in a claimed
XX method for making recombinant matrilysin. The zymogen (inactive)
XX form of matrilysin is a single-chain protein. The active 2-chain
XX form strongly interacts with fragments of a Kunitz-type serine
XX protease inhibitor (hepatocyte growth factor activator inhibitor,
XX HAI-1) to form SDS-stable complexes. In breast cancer cells,
XX matrilysin is present mainly as the uncomplexed form. Only the
XX complexed matrilysin is detected in human milk. The invention is
XX directed to a method of detecting a malignancy or a pre-malignant
XX lesion in breast or other tissue by detecting the presence of
XX single- or 2-chain forms of matrilysin in the tissue. The
XX object is to inhibit tumour onset, tumour growth and metastasis.
XX Malignancies and pre-malignant conditions characterized by
XX expression of the zymogen or activated form of matrilysin are
XX treated by administering an inhibitor of matrilysin, especially a
XX Bowman-Birk inhibitor. The pre-malignant condition is atypical
XX ductal hyperplasia of the breast, actinic keratosis, leukoplakia,
XX Barrett's epithelium of the esophagus, ulcerative colitis,
XX adenomatous colorectal polyps, erythroplasia of the Queyrat,
XX Bowen's disease, Bowenoid papulosis, vulvar intraepithelial
XX neoplasia or dysplastic changes to the cervix. The invention also
XX provides methods for in vivo or in vitro diagnosis of malignancy
XX or pre-malignant lesion, and methods of identifying matrilysin
XX modulators, including activators and inhibitors.
XX
XX Sequence 2955 BP; 602 A; 909 C; 885 G; 559 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 4,5e-30 Length: 2955
XX Score: 488.00 Matches: 186
XX Percent Similarity: 37.81% Conservative: 90
XX Best Local Similarity: 25.488 Mismatches: 270
XX Query Match: 12.99% Indels: 184
XX DB: 21 Gaps: 31
XX
XX US-09-874-198-2 (1-686) x AAA88492 (1-2955)
QY 29 ArgLeuAlaSerProGlyPhePro---GlyGluTyrAlaAsnAspGlnIleuArgArgTrp 47
DB 520 CGCTTCACACCGCCGCTTCCTCCAGACCCCTACATCCGCTTCGCGAGTGG 579
QY 48 ThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisIleAspLeuGlu 67
DB 580 GGCCTCGGGGGGAGCCGACATCAGCTGAGCTTCACCTCCGCGAGCTTGTGCG 639
QY 68 LeuSerHisLeuGlyGluTyrAspPheValLysLeuSerSerGlyValAlaLys----- 84
DB 640 TCGTCGACGACGCGCGACGACGCTGTGAGGTGTACACACCCGTGAGCCCATGAG 699
QY 85 -----ValLeuAlaThrLeuGlyGlnGluSerThrAspThrGlnArgAlaProGly 102
DB 700 CCCACGCGCCCTGTGTCAGTTGTGTGC-----ACATACCCCTCCCTCTAC 744

```

OY	103	LysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerLysPyrSer	122
Db	745	AACGTGCATCTTCAGCTCCCTCCAGAACGTCCTGCTCATCACTGTAAACAACGTAG	804
OY	123	AsnGluLysProPheThrGlyPheGluAla---PheTyralaIaGluAspIleaSgu	141
Db	805	CGCGGGCATTCCC-----GGCTTTGAGGCCACTTCTCCAGCTGGCTTAGATGAACAG	858
OY	142	CysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisasnhIsleuGly	161
Db	859	TGT-----	861
OY	162	GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrcysSer	181
Db	861	-----	861
OY	182	AlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyCLeuSerSerProGluTyr	201
Db	862	-----GGAGCCCGCTTAAGTAAAGCCAGGGGACATTACAGCCCCCTACTAC	909
OY	202	ProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGlyPhe	221
Db	910	CGAGGCCAATCCACCACCAACATGATGATGCATGACATGAACTTGAGTGCCCAACAACAG	969
OY	222	SerValIleLeuAspPheValGluSerPheAspValGluThrHis---ProGluThrLeu	240
Db	970	CATGTGAAGTGGCCCTCAATCTTCTTCACTCTGAGGCCGCGCTCGGACACC	1029
OY	241	CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnIslaGlyProPheCysGly	260
Db	1030	TGCCCCAAGCACTACGTGGAAATCAATGGGAGAATA-----TACTGCGGA	1074
OY	261	LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr	280
Db	1075	GAGAGTCCACGATTCGTCGTCACAGACAGACAGACAAAGATCACTAGCTTCCACATA	1133
OY	281	AspGluSerGlyAspHisThrGlyTyrPylsIleHisTyrThrSer-----ThrAlaGln	298
Db	1135	GATCAGTCTCTACACCGACACCGCGCTTCTTAGCTGAATACCTCTCCACATCCACAGTAC	1194
OY	299	ProCysProTyrPrometAlaPropromsnglyHisValserProvalImLaIalysTyr	318
Db	1135	CCATGCCCG-----GGCACGTTCACGTGCCGACGGGGGGGTGT	1233
OY	319	IleLeuLys-----AspSerPhe	324
Db	1234	ATCCGGAAGAGAGCGCCCTGATGCTGGGGCCGACTGCACGACGACCAKAGCATGAGCTC	1299
OY	335	SerIlePheCysgluThrGlyTyrGlnLeuLeuGlnGlyHisIleuProLeuLysSerPhe	344
Db	294	AACGTCAATTCGACGCGCGGCCACACAGTTCACGTGCAGAACAAAGTTGTGCAAAGCCCTTC	1355
OY	345	ThrAlaValCysGlnLys-----AspGlySerTyrAspArgPrometPro	359
Db	1354	TTCTGGGCTCTCGACAGTGTGAACGACTGCGGAGACAGACAGCAGCAGCAGCAG-	1404
OY	360	AlaCysSerIleVal-----	364
Db	1405	GGGTGCACATTTCCGGCCGACAGCTTCAGGTTCACATGGGAAGAGCTCTTCGCAAAAAC	1466
OY	365	-----AspCysGlyPro---ProAspAspLeuProSerGlyArg	376
Db	1465	CAGCAGTGCATGAGGAGAGACAGACTGGGGGAGCGGAGCGAGAGGCTCTCGGCCCAAG	1522
OY	377	ValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCys	396
Db	1535	GTGAACGTGCACCTTGTACCAACACACC-----TACGCGTGC	1563
OY	397	GluGluThrPheTyrThrMetLysValAsn-----AspGlyLysTyrValCysGlu	413
Db	1564	CTCAATGGGCTCTGCTGAGCAAGAGGCAACCCCTGAAGTGTGACGGGAAGGAGCAGTGT--	1622
OY	414	AlaAspGlyPheTrpThrSerSerLysGlyGluLysSerLeuProValCysGluProVal	433

Db	1621	AGCGACGGC-----TCAGATGAAACAGAC-----TGCAC-----	1650
Qy	434	CysglLeuSerAlaArgThrThrGlyGlyArgIleTyrGlyGlyGlnLysAlaLysPro	453
Db	1651	TTGTGGCGTGGCGGTATTCACGACACAGCGCTCGTGTGGGGGACGAGTCCGGATAG	1710
Qy	454	GlyAspPheProTrpGlnValLeuIle----LeuGly---GlyThrAlaAlaGly	470
Db	1711	GGCGAGTGGCCCTGGCAGAGTAAGCGTCATGCTCTGGCCAGGCGCATCTGGCGTCT	1770
Qy	471	AlaLeuLeuTyrAspAsnTrpValLeuLeuThrAlaAlaHisAlaValTyrGlnLys---	489
Db	1771	TCCCTCATCTCTCCCACTGGCTGGTCTGTGGCGGACGCTACATGATGACAGAGA	1830
Qy	490	-----HisAsp-----	491
Db	1831	TTTCAGGTACTACAGACCCACGACGTGAGCGGCTTCCTGGGCTTGACAGCACAGACAG	1890
Qy	492	AlaSerAlaLeuAsnProIleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGln	511
Db	1891	CGCAGCGCCCGCTGGGGGTGACGAGAGCGGACGCTCAGACGATCATCTCCAC	1941
Qy	512	AlaTrpSerGluAlaValPheIleHisGluGlyTyrThrHisAspAlaGlyPheAspAsn	531
Db	1942	-----CCCTCTTCAAAGACTTCACCTTCGCACAT	1971
Qy	532	AspIleAlaLeuIleLysLeuAsnAsnLysValValIleAsnSerAsnIleThrProIle	551
Db	1972	CACATCGCGGCTGAGAGTGGAGTGGAGAAACCGGACGAGTACAGCTCCATGGTGGGCCCATC	2031
Qy	552	CysLeuProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSer	571
Db	2032	TGCGTCGGC-----GACGGCTCCCATGTCCTTCCCTGGCGGACAGCCATCTGGGTACG	2085
Qy	572	GlyTrpGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetTyrValAspIle	590
Db	2086	GGCTGGGGACACACCCAGTATGGAGCGACCTGGCGGCTGATCTCCAAAGGCTGCAGATC	2145
Qy	591	ProIleValAspHisGlnLysCysThrAlaAlaTyrGlnLysProProTyrProArgGly	610
Db	2146	CGCGTATCAACACACACCCACTGGAGAACCTCTGCCGAC-----	2187
Qy	611	SerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGlyLysAspSerCysArg	630
Db	2188	CAGATCACAGCCCGCATGATGTGCGTGGCTTCTCTCAGGGGCGGTGACATCTGCCACG	2247
Qy	631	GlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGluArgTrpPheValGly	650
Db	2248	GGTGATGTCCGGGGGACCCCTGTCACAGCTGGAGCGCGAT---GGCGGATCTTCCAGGCC	2304
Qy	651	GlyIleValaSerTrpGlySerMetAsnGlyGluAlaGlyGlnTyrGlyValTyrThr	670
Db	2305	GGTGGTGGAGACTGGGAGAC---GGCTGGCGTCAAGAGAACACCGCGGTGTACACA	2361
Qy	671	LysValIleAsnTyrIleProTrpIleGlu	680
Db	2362	AGGCTCCCTCTGTTCGGGAGCTGGATCAAA	2391
RESULT 24			
AAAF28099			
ID	AAF28099 standard; cDNA; 3142 BP.		
AC	AAF28099;		
XX	06-JUN-2001 (first entry)		
DT			
XX	Human membrane-type serine protease MT-SPL coding sequence.		
DE	Human membrane-type serine protease; MT-SPL; cancer; ss.		
KN			
XX	Homo sapiens.		
OS			
XX			


```

Db 1966 TCCCTCATCTTCCCACTGGCTGTTCTCTCGCGACACTGCTACATGCATGACAGAGA 2025
OY 490 -----HisASP----- 491
Db 2026 TTCAGTACTCAGACCCCAACCCAGTGCAGCGCTTCTCGGCTTGCCAGCCAGAGCCAG 2085
OY 492 AlaserAlaLeuaspIleArgMetGlyThrLeuLysArgLeuSerProHisThrGln 511
Db 2086 CGCAGCGCCCTGGGTGGGAGGAGCGCAGAGCTCAAGCGCATCTCTCCAC----- 2136
OY 512 AlatrpserIuAlaValPheIleHisGluGlyThrHisAspAlaGlyPheAspAsn 531
Db 2137 -----CCTCTTCTCAATGACTTCCACCTTGACTAT 2166
OY 532 AsplleAlaLeuIleLysLeuAsnLysValValIleAsnSerAsnIleThrProIle 551
Db 2167 GACATCGCGCTGCTGGAGCTGGAGAAACCGGCAGAGTACAGCTTCATGATGCGGCCCATC 2226
OY 552 CysLeuProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSer 571
Db 2227 TGCCTGCCG-----GACGCCCTCCCATGCTCTCCCTCGCGCAAGGCCATCTGGGTACG 2280
OY 572 GlyTrpGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetTyrrValAspIle 590
Db 2281 GGCTGGGACACACCCAGTATGAGGCACCTGGCGCGCTGATCTGCAAAAGGAGATC 2340
OY 591 ProIleValAspHisGlnLysCysThrAlaAlaTyrrGluLysProProTyrrProArgGly 610
Db 2341 CGGCTCTACACACAGACCACTGCGAGACCTCTCGCCGAG----- 2382
OY 611 SerValThrAlaAsnMetLeuLysAlaGlyLeuGluSerGlyLysAspSerCysArg 630
Db 2383 CAGATCAGCGCGCATGATGCTGGTGGCTCTCTCAGCGCGCGCATCTCTCGCCAG 2442
OY 631 GlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGlnArgTrpPheValGly 650
Db 2443 GGTGATTCCGGGGACCCCTGTCACGCTGAGAGCGGAT--GGCGCATCTTCCAGGCC 2499
OY 651 GlyIleValSerTrpGlySerMetAsnCysGlyGluAlaGlyGlnTyrrGlyValTyrrThr 670
Db 2500 GGTGTGTGAGCTGGGGAGAC---GGCTGGCGTACAGAGAACAGCAGCGGTGTACACA 2556
OY 671 LysValIleAsnTyrrIleProTrpIleGlu 680
Db 2557 AGGCTCCCTCTGTTTCGAGCTGATCAAA 2586
RESULT 25
AAA88493
AAA88493; standard; cDNA; 3149 BP.
AAA88493;
22-JAN-2001 (first entry)
XX DE Human matrilipase cDNA.
XX DE
XX DE Matrilipase; serine protease; human; breast cancer; pre-malignancy;
XX DE actinic keratosis; leukoplakia; Barrett's epithelium;
XX DE columnar metaplasia; ulcerative colitis; Bowenoid papulosis;
XX DE adenomatous colorectal polyp; Ovarian epithelioma;
XX DE vulvar intraepithelial neoplasia; tumour; metastasis; therapy; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT CDS 36..2603
XX FT /*tag= c
XX
XX MO200053232-A1.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000MO-US06111.

```

```

XX PR 12-MAR-1999; 99US-0124006.
XX PA (GEOU ) UNIV GEORGETOWN.
XX PI Dickson RB, Lin C, Johnson M, Wang S, Emyedy I;
XX DR MPI: 2000-594268/56.
XX DR P-PSDB; AAB19552.
XX
XX PT Treating malignancies, premalignant and pathologic conditions in a
XX PT subject, comprises administering matrilipase modulating agent
XX PS Claim 9; Fig 15; 116pp; English.
XX
XX The present sequence is that of cDNA coding for the full-length
XX form (see AAB19521) of human matrilipase, a trypsin-like protease.
XX cDNA encoding a truncated form of matrilipase is given in AAA88492.
XX Either form can be incorporated into a vector and used in a claimed
XX method for making recombinant matrilipase. The zymogen (inactive)
XX form of matrilipase is a single-chain protein. The active 2-chain
XX form strongly interacts with fragments of a kunitz-type serine
XX protease inhibitor (hepatocyte growth factor activator inhibitor,
XX HAI-1) to form SDS-stable complexes. In breast cancer cells,
XX matrilipase is present mainly as the uncomplexed form. Only the
XX complexed matrilipase is detected in human milk. The invention is
XX directed to a method of detecting a malignancy or a pre-malignant
XX lesion in breast or other tissue by detecting the presence of
XX single- or 2-chain forms of matrilipase in the tissue. The
XX object is to inhibit tumour onset, tumour growth and metastasis.
XX Matrilipases and pre-malignant conditions characterised by
XX expression of the zymogen or activated form of matrilipase are
XX treated by administering an inhibitor of matrilipase, especially a
XX Bowman-Birk inhibitor. The pre-malignant condition is atypical
XX ductal hyperplasia or the breast, actinic keratosis, leukoplakia,
XX Barrett's epithelium of the esophagus, ulcerative colitis,
XX adenomatous colorectal polyps, erythroplasia of the cervix,
XX Bowen's disease, Bowenoid papulosis, vulvar intraepithelial
XX neoplasia or dysplastic changes to the cervix. The invention also
XX provides methods for in vivo or in vitro diagnosis of malignancy
XX or pre-malignant lesion, and methods of identifying matrilipase
XX modulators, including activators and inhibitors.
XX
XX Sequence 3149 BP; 647 A; 959 C; 962 G; 581 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 4,89e-30 Length: 3149
Score: 488.00 Matches: 186
Percent Similarity: 37.81% Conservative: 90
Best Local Similarity: 25.48% Mismatches: 270
Query Match: 12.99% Indels: 184
DB: 21 Gaps: 31
US-09-874-198-2 (1-686) x AAA88493 (1-3149)
OY 29 ArgLeuAlaSerProGlyPhePro---GlyGluTyrrAlaAsnAspGlnArgArgTrp 47
Db 714 CGTTTACACACAGCGCGCTTCCCTGACAGCCCTTACCCCGCTGATCCCGGCTGCAGTGG 773
OY 48 ThrLeuThrAlaProProGlyTyrrArgLeuValGlyLeuTyrrPheThrHisPheAspLeuGlu 67
Db 774 GCCCTGCGGGGAGCCCGCATGCTGAGCTGACCTCCAGCTTCCGACGCTTTCGCG 833
OY 68 LeuSerHisLeuLysGluTyrrAspPheValLysLeuSerSerGlyAlaLys----- 84
Db 834 TCCTGGACAGCGCGCAGCAGCTGTGACGCTGTACAAACCCCTGAGCCCATGAG 893
OY 85 -----ValLeuAlaThrLeuLysGlyGlnGluSerThrAspThrGlnArgAlaProGly 102
Db 894 CCCACGCCCTGTGTGACAGTTGTGTGC-----ACCTAACCTCCCTCTTAC 938
OY 103 LysAspThrPheTyrrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrrSer 122

```

Db 939 AACCTGACCTTCACGTCCTCCAGAACGTCCTCATCATCACTGATACCAACACTGAG 998
 QY 123 ASngLdysProPheThrGlyPheGluAla---PheTyrAlaAlaGluAspIleAspGlu 141
 Db 999 GGGGGGATCC-----GGCTTTGAGGGCACCCTTCTCCAGCTGCTCAGATGAGCAGC 1052
 QY 142 CysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGly 161
 Db 1053 TGT----- 1055
 QY 162 GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSer 161
 Db 1055 ----- 1055
 QY 182 AlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyLeuSerSerProGlyTyr 201
 Db 1056 -----GGAGGCCCTTACGTAAGGCCAGGGGACATTCAACAGCCCTTACTAC 1103
 QY 202 ProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGlnGlyPhe 221
 Db 1104 CCAGGCGACATACCACCAACATTTGACTGCACATGGAACATTGAGGTGCCCAACACAG 1163
 QY 222 SerValIleLeuAspPheValIleGluSerPheAspValIleThrHis---ProGluThrLeu 240
 Db 1164 CATGTAAAGGTGGCTTCAAAATCTTCTACCTGTGAGCGCGCTGCGGGGAC 1223
 QY 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGlnGlnHisGlyProPheCysGly 260
 Db 1224 TGCCCAAGACACTACGGGATCATGGGACAAA-----TACTGGGA 1268
 QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr 280
 Db 1269 GAGAGGTCCAGTTCGTCGACACACACAGCAACAGTCAAGTTCAGTCTTCACTCA 1328
 QY 281 AspGlnSerGlyAspHisThrGlyTyrLysIleHisTyrThrSer-----ThrAlaGln 298
 Db 1329 GATCAGTCCACACCCGACCGGCTTCTTACGTGAATCCTCTCAACGCTCCAGTAC 1388
 QY 299 ProCysProTyrProMetAlaProProAsnGlnHisValSerProValGlnAlaLysTyr 318
 Db 1389 CCATGGCCG-----GGGACGTTCAACGTCGCGCCGACGGGGCGGTGT 1427
 QY 319 IleLeuLys-----AspSerPhe 324
 Db 1428 ATCCGAGAGAGCTGCGCTGTGATGGCTGGCGACTGCACCCAGCACAGCGATGAGCTC 1487
 QY 325 SerIlePheCysGlnThrGlyTyrGluLeuLeuGlnGlyHisIleLeuProLeuLysSerPhe 344
 Db 488 AACTGCAGTTGCGACGCGCGCCAGCTTCACTGCAGAGAACAGTTCTGCAAGCCCTTC 1547
 QY 345 ThrAlaValCysGlnLys-----AspGlySerThrPaspArgPromePro 359
 Db 1548 TTCTGGGTGCGACAGTGTGAACGACTGGGAGACACAGCAGCAGCAG----- 1598
 QY 360 AlaCysSerIleVal----- 364
 Db 1599 GGGTGCAGTGTGCGGCCAGACCTTCAGGTGTTCATGGAAGTGCCTCTCGAAGAGC 1658
 QY 365 -----AspCysGlyPro---ProAspAspLeuProSerGlyArg 376
 Db 1659 CAGCACTGCAGTGGAGAGCAGGAGCGGCTCCGAGGCGCTCTGCTCCGCAAG 1718
 QY 377 ValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCys 396
 Db 1719 GTGAAGCTGTACTGTGACCAACACACC-----TACCGCTGC 1757
 QY 397 GluGluThrPheTyrThrMetLysValAsn-----AspGlyLysTyrValCysGlu 413
 Db 1758 CTCGAATGGCTCTGTTGAGCAGGCAACCTGAGTGTGACGGAAGAGGACTGT--- 1814
 QY 414 AlaAspGlyPheThrPheThrSerLysGlyGluLysSerLeuProValCysGluProVal 433
 Db 1815 AGCGACGGC-----TCAGATGAGAGAGAC-----TCCGAC----- 1844

QY 434 CysGlyLeuSerAlaArgThrThrGlyArgIleTyrGlyGlnLysAlaLysPro 453
 Db 1845 TGTGGGTGGCGGTCATTCACAGACAGCGCTGCTGTTGTTGGGGGACAGGATGGGATGAG 1904
 QY 454 GlyAspPheProThrGlnValLeuIle-----LeuGly---GlyThrThrAlaAlaGly 470
 Db 1905 GGGCAGTGGCCCTGGCGAGTAAGCTTCATGCTGTGGGCCACGGCCATCTCGGTGT 1964
 QY 471 AlaLeuLeuTyrAspAsnThrPheValLeuThrAlaAlaHisAlaValTyrGluGlnLys--- 489
 Db 1965 TCCCTCATCTCCCACTGGCTGTGCTTCCGACACATGTAATGATGACAGAGA 2024
 QY 490 -----HisAsp----- 491
 Db 2025 TTCAGGTACTCAGACCCACGACGTGAGCGCTTCTGCGGCTTGCAGCAGAGCCAG 2084
 QY 492 AlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGln 511
 Db 2085 GCGAGCGCCCTGGGTGCGAGCGCAGCGCTCAAGCGCATCTTCCAC----- 2135
 QY 512 AlaThrSerLeuAlaValPheIleHisGluGlyTyrThrHisAspAlaGlyPheAspAsn 531
 Db 2136 -----CCCTTCTCAATGACTTCACTTCGACTAT 2165
 QY 532 AspIleAlaLeuIleLysLeuAsnAsnLysValIleAsnSerAsnIleThrProIle 551
 Db 2166 GACATCCGCGCTGCTGAGCTGAGAAACCGCAGAGTACAGCTTCATGATGCGGCCATC 2225
 QY 552 CysLeuProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSer 571
 Db 2226 TGCCCTGCG-----GACGCTCCCATGTCTCTCCGCGCAAGCGCATCTGGGTACG 2279
 QY 572 GlyThrGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnMetLysValAspIle 590
 Db 2280 GGTGGGGACACACCCAGTATGAGGCACTGGCGCGCTGATCTTGCAGAAAGGTGAGATC 2339
 QY 591 ProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProProTyrProArgGly 610
 Db 2340 GCGGTATTCACACAGACCACTGCGAGAACCTCTCGCGCAG----- 2381
 QY 611 SerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyLysAspSerCysArg 630
 Db 2382 CAGATCACGCGCGCATATGATGCTGCGCTCTCTCCAGCGCGCGTCTCTGCCAG 2441
 QY 631 GlyAspSerGlyGlyAlaLeuValPheLeuAspSerCylThrGluArgThrPheValGly 650
 Db 2442 GGTGATTCGCGGGACCCCTGTCCAGCGTGAAGCGGAT---GGGCGATCTTCCAGGCC 2498
 QY 651 GlyIleValSerThrPglYserMetAsnCysGlyGluAlaGlyGlnTyrGlyValTyrThr 670
 Db 2499 GGTGTGTGAGCTGGGAGAC---GGCTGGCTCAGAGGAACAAGCAGGCTGTACACA 2555
 QY 671 LysValIleAsnTyrIleProThrIleGlu 680
 Db 2556 AGGCTCCCTCTGTTTGGAGCTGATCAAA 2585

Search completed: January 23, 2003, 22:32:16
 Job time : 415 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 23, 2003, 22:24:46 ; Search time 79 Seconds
(without alignments)
2663.041 Million cell updates/sec

Title: US-09-874-198-2
Perfect score: 3758
Sequence: 1 MRLTLGLGSGVATPLGP.....GYTKVINYIPWNIISDF 686

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Sequences: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 75 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=xlp
-CGN2_1/USPTO.spool/US09874198/runat_22012003_085001_10913/app.query.fasta_1.839
-DB=Issued_Patents.NA -QWRT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPL=0
-DOOPL=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=75 -DOCALLGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=25
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09874198.ecgn1.1.31.6runat.22012003_085001_10913 -NCP=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEMEMORY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-GAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents.NA.*

- 1: /cgn2_6/ptodata/1/lna/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/lna/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/lna/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/1/lna/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/1/lna/PCRTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/1/lna/Deckfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1412	37.6	2492	4	US-09-381-779-1
2	490	13.0	3147	4	US-09-027-337-1
3	490	13.0	3147	4	US-09-644-600-1
4	490	13.0	3147	4	US-09-644-600-18
5	470.5	12.5	3448	1	US-08-296-014A-3
6	470.5	12.5	3448	1	US-08-296-014A-1
7	470.5	12.5	3448	2	US-08-596-405-3
8	470.5	12.5	3448	2	US-08-877-620-3
9	470.5	12.5	4182	1	US-08-296-014A-1
10	470.5	12.5	4182	2	US-08-596-405-1
11	446.5	11.9	4182	2	US-08-877-620-1
12	432.5	11.5	807	4	US-09-715-994-1
			2900	2	US-09-027-337-9

13	432.5	11.5	2900	4	US-09-644-600-9	Sequence 9, Appl1
14	411	10.9	3919	2	US-08-866-650-4	Sequence 4, Appl1
15	411	10.9	3919	2	US-09-021-287-4	Sequence 4, Appl1
16	411	10.9	3919	4	US-09-240-473-4	Sequence 4, Appl1
17	411	10.9	5145	3	US-08-991-408-1	Sequence 1, Appl1
18	411	10.9	5145	3	US-09-432-473-1	Sequence 1, Appl1
19	410	10.9	3690	4	US-08-991-408-3	Sequence 3, Appl1
20	410	10.9	3690	4	US-09-432-473-3	Sequence 3, Appl1
21	408.5	10.9	2581	1	US-08-200-900A-1	Sequence 1, Appl1
22	408.5	10.9	2581	5	PCR-US84-00616-1	Sequence 1, Appl1
23	405	10.8	1389	4	US-09-202-101-16	Sequence 16, Appl1
24	405	10.8	1389	6	5521070-1	APPLICANT: MEU
25	405	10.8	1548	1	US-08-106-078-15	Sequence 15, Appl1
26	405	10.8	1548	1	US-08-591-492-21	Sequence 21, Appl1
27	405	10.8	1548	2	US-08-484-891-6	Sequence 6, Appl1
28	405	10.8	1510	1	US-08-209-845A-6	Sequence 6, Appl1
29	405	10.8	1610	2	US-08-472-809B-6	Sequence 6, Appl1
30	405	10.8	1610	4	US-08-438-265-6	Sequence 6, Appl1
31	405	10.8	2775	4	US-09-053-871A-22	Sequence 22, Appl1
32	405	10.8	2802	3	US-08-742-877-1	Sequence 1, Appl1
33	405	10.8	5532	3	US-08-651-472-72	Sequence 72, Appl1
34	405	10.8	5532	4	US-08-358-928-72	Sequence 72, Appl1
35	403.5	10.7	3546	4	US-08-872-757-3	Sequence 3, Appl1
36	401.5	10.7	1437	2	US-08-479-733A-26	Sequence 26, Appl1
37	401.5	10.7	1437	3	US-08-487-427-26	Sequence 26, Appl1
38	401.5	10.7	1437	3	US-08-479-727A-26	Sequence 26, Appl1
39	401.5	10.7	1437	3	US-08-482-369A-26	Sequence 26, Appl1
40	401.5	10.7	1437	5	PCR-US85-07439-26	Sequence 26, Appl1
41	401	10.7	2487	1	US-08-377-292-1	Sequence 26, Appl1
42	398.5	10.6	1404	4	US-09-202-101-15	Sequence 15, Appl1
43	398	10.6	1605	2	US-09-000-846-1	Sequence 1, Appl1
44	395.5	10.5	1783	4	US-09-510-738A-188	Sequence 188, App
45	395	10.5	2363	4	US-08-742-703-3	Sequence 3, Appl1
46	394	10.5	2457	4	US-08-872-757-1	Sequence 1, Appl1
47	393.5	10.5	4771	2	US-08-866-650-2	Sequence 2, Appl1
48	393.5	10.5	4771	2	US-09-021-287-2	Sequence 2, Appl1
49	393.5	10.5	4771	2	US-09-240-473-2	Sequence 2, Appl1
50	386.5	10.3	1126	2	US-08-479-733A-27	Sequence 27, Appl1
51	386.5	10.3	1126	3	US-08-487-427-27	Sequence 27, Appl1
52	386.5	10.3	1126	3	US-08-479-722A-27	Sequence 27, Appl1
53	386.5	10.3	1126	3	US-08-482-369A-27	Sequence 27, Appl1
54	386.5	10.3	1126	5	PCR-US85-07439-27	Sequence 27, Appl1
55	384	10.2	1869	3	US-08-952-967-7	Sequence 7, Appl1
56	382.5	10.2	1386	2	US-08-756-506-3	Sequence 3, Appl1
57	382.5	10.2	1755	6	5225537-1	Patent No. 5225537
58	379	10.1	1947	1	US-07-998-972A-2	Sequence 2, Appl1
59	379	10.1	1947	1	US-08-463-953-2	Sequence 2, Appl1
60	379	10.1	1947	2	US-08-462-261-2	Sequence 2, Appl1
61	379	10.1	1947	2	US-08-479-733A-24	Sequence 24, Appl1
62	379	10.1	1947	3	US-08-487-427-24	Sequence 24, Appl1
63	379	10.1	1947	3	US-08-479-722A-24	Sequence 24, Appl1
64	379	10.1	1947	3	US-08-482-369A-24	Sequence 24, Appl1
65	379	10.1	1947	5	PCR-US92-11357-2	Sequence 2, Appl1
66	379	10.1	1947	5	PCR-US95-07439-24	Sequence 24, Appl1
67	379	10.1	1988	1	US-07-750-080A-15	Sequence 15, Appl1
68	379	10.1	1988	3	US-08-651-472-15	Sequence 15, Appl1
69	379	10.1	1988	4	US-08-358-928-15	Sequence 15, Appl1
70	376.5	10.0	1387	6	5270178-1	Patent No. 5270178
71	371.5	9.9	1894	4	US-09-004-731-29	Sequence 29, Appl1
72	371.5	9.9	1894	4	US-09-004-731-31	Sequence 31, Appl1
73	371.5	9.9	1894	4	US-09-032-215-3	Sequence 3, Appl1
74	371.5	9.9	1894	4	US-08-749-699-29	Sequence 29, Appl1
75	371.5	9.9	1894	4	US-08-749-699-31	Sequence 31, Appl1

ALIGNMENTS

RESULT 1
US-09-381-779-1
; Sequence 1, Application US/09381779
; Patent No. 6190868
; GENERAL INFORMATION:
; APPLICANT: Rothberg, Jonathan M.

APPLICANT: Deem, Michael W.
APPLICANT: Simpson, John W.
TITLE OF INVENTION: METHOD FOR IDENTIFYING A NUCLEIC ACID SEQUENCE
FILE REFERENCE: Cura-8 US 15966-508
CURRENT APPLICATION NUMBER: US/09/381,779
CURRENT FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: U.S.S.N. 60/054,887
PRIOR FILING DATE: 1997-08-07
PRIOR APPLICATION NUMBER: PCT/US98/16548
PRIOR FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 1
LENGTH: 2492
TYPE: DNA
ORGANISM: Homo sapiens
US-09-381-779-1

Alignment Scores:

No.:	1,32e-135	Length:	2492
Global Similarity:	1412.00	Matches:	301
Local Similarity:	56.73%	Conservative:	108
Query Match:	41.75%	Mismatches:	248
DB:	37.57%	Indels:	64
	4	Gaps:	22

US-09-874-198-2 (1-686) x US-09-381-779-1 (1-2492)

QY 1 MetArgLeuLeuThrLeuLeu-----GlyLeuLeuCys-----GlySerValAla 15
DB 64 ArgTGCTGCTTGTACCTGCTGCGCGCCGCTGTCGACGGGACGAGGAGCTTCATTT--- 120
QY 16 ThrProLeuGlyProLysTrpProLysProValPheGlyArgLeuAlaSerProGlyPhe 35
DB 121 -----CCATC-----CCTCAGAGTTATTGCGGAGGTGACTTCCCTGCTTC 165
QY 36 ProGlyGluTyrAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 55
DB 166 CCACACCTTACCCCAACAATTGAAACACACCTGTGATCCAGTCCGACGGGATAC 225
QY 56 ArgLeuArgLeuTyrPheThrHisPheAspLeuGluLeuSerHisLeuGlySerGlyTyrAsp 75
DB 226 AGGGTAAAGCTGCTGTCCAGCAGTTTGACCTGTGAGCCCTTGTGAAGCTGCTTCTATGAT 285
QY 76 PheValLeuLeuSerSerGlyAlaValLeuAlaThrLeuGlyGlnGlnSerThr 95
DB 286 TATGTCAAGATCTCTGCTGATAGAAAGCCTGGGAGGTCTGTGCGCAACGCTGCTCT 345
QY 96 AspThrGluArgAlaProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIle 115
DB 346 CCACGTGGGCAACCCCGGGAAGAAAGAAATTTATGTCACAAAGGAAGATGCTGCTG 405
QY 116 ThrPheArgSerAspTrpSerAsnGluLysPro-----PheThrArgPhe 130
DB 406 ACCCTTCACACACAGCTTCTCAACAGGAGAAATGGACATCATCTGTTTACAAAGGCTTC 465
QY 131 GluAlaPheTyrAlaAlaGluAspIleAspGluCys-----GlnValAlaProGlyLys 148
DB 466 CTGCGCTACTACCAACCTGTGACCTTGATGATGCTTCGCCGAGCAAAATAGGGGAG 525
QY 149 -----AlaProThrCysAspHisHisCysHisAsnHisLeuGlyLeuGlyPheTyrCys 165
DB 526 GAGGATCCCGACGCGCCAGTGCAGCAGCTGTGCACAACTAGCTTGAGGCTACTCTG 585
QY 166 SerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSer 185
DB 586 TCCTGCGCTACAGGCTATAGGCTTCAGAGAGACAGACAGCATTCGCAAGGCTGATGCACAG 645
QY 186 GlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGlyTyrProArgProTyr 205
DB 646 AGCGACCTGTACACGAGGATCAGGCTACATCTCCAGCCTGGAGTACCTCGGCTCTAC 705
QY 206 ProLysLeuSerSerCysThrTyrSerIleSerLeuGluGluGlyPheSerValIleLeu 225

DB 706 CCCCGTAGCTCGCTGCACTACAGACATCCGGGGAGCGGGGCTCAACCTGCACCTC 765
QY 226 AspPheValGluSerPheAspValGluThrHisProGlyThrLeuCysProTyrAspPhe 245
DB 766 AAGTTCCTGGAGCTTTTATATGTGACCCAGCAAGTACTGCTGCTTATGACCG 825
QY 246 LeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGlyLysThrLeuProHis 265
DB 826 CTACAGATCATTCACAGGAAACATTTGGAGTTCTGGGAGAGCAAGGCCCCC 885
QY 266 ArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGluSerGlyAsp 285
DB 886 GACCTCGACACAGCAGCATCTGTGATCTGCTTTCACAGATAGTCGGGGGC 945
QY 286 HisThrGlyTrpLysIleHisThrSerThrAlaGlnProCysProTyrProMetAla 305
DB 946 ACCGGGCTGGAGAGCTGCGCTACACCCAGATCATCAAGTGGCCCGACCAAGACC 1005
QY 306 ProProAsn-----GlyHisValSerProValGlnAlaLysTyrIleLeuLysAspSerPhe 324
DB 1006 CTAGACGATTCACCATCATTCAGAACTGCAGACCTGCATACAGTTCGCTGACTTTC 1065
QY 325 SerIlePheCysGluThrGlyTyrGluLeuGlnGlyHisLeuProLeuLysSerPhe 344
DB 1066 ATTGCTACTCTGCACAGCAGTACACAGCTCATAGAGGAGGACAGGCTGCTGCTTTC 1125
QY 345 ThrAlaValCysGlnLysAspGlySerTrpAspArgProMetProAlaCysSerIleVal 364
DB 1126 ACAGCTGTCTGCAGATATATGACCGTGACATCGCGCTCCACGATCCACAGATCAAG 1185
QY 365 AspCysGlyProProAspAspLeuProSerGlyArgValGlyTyrIleThrGlyProGly 384
DB 1186 GACTGTGGGACCGCCGAGATGCTGCTATGATGATGACTTCCGTACACCAACCAATGGA 1245
QY 385 ValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMetLys 404
DB 1246 GTGACACCTACAGAGCCCGTATCCAGTACTGCTGCAATGACCAATATCAAGATGACG 1305
QY 405 Val-----AsnAspGlyLysTyrValCysGluLysAspGlyPhe 417
DB 1306 ACCAGCTGCGCAGCAGGAGCTGTGACGAAAGGGTGTACACTGCGACACAGGCAATT 1365
QY 418 Trp---ThrSerSerLysGluLysSerLeuProValCysGluProValCysGly--- 435
DB 1366 TGGAAATAGAACAGAGAGGAGAG---ATTCTCGGCTGCTGCACTGTGTGGAGAG 1422
QY 436 ---LeuSerAlaArgThrThrGlyAlaGlyIleTyrGlyGlyGlnLysAlaLysProGly 454
DB 1423 CCCGTGAACCCCGTGAACAGAGGAGCGCATATCTGGAGGCGCAAAACCAAGATGGGC 1482
QY 455 AspPheProTyrGlnValLeuIleLeuGlyGlyThrThrAlaAlaGlyLysLeuTyr 474
DB 1483 AACTTCCCTGCGAGGTGTTCACCAACATCCACAGGCGGGGGCGGGCCCTGCTGGGC 1542
QY 475 AspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysHisAspAla----- 492
DB 1543 GACCGGTGATCTGCACACTGCGCACCTGTATCCCAAGAACACAGACGCAAAACC 1602
QY 493 ---SerAlaLeuAspIleArgMetGly-----ThrLeuLysArgLeuSer 506
DB 1603 AACGCTCTTTGATGCTGTCTGCGCCACACAAATGTGAAGAGCTCATGAGCTAGGA 1662
QY 507 ProHisTyrThrGlnAlaTrpSerGluAlaValPheIleHisGluGlyLysThrHisAsp 526
DB 1663 AATCACCCCATCCGC-----ACGGTACGCGTCAACCCGAGCTACCGTACGAGAT 1710
QY 527 Ala-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValIle 544
DB 1711 GAGTCTACAAATTTTGAAGGGGACATCGCGCTGTGAGACTGGAAATATGTGTACACCTG 1770
QY 545 AsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThr 564

```
Db 1771 GGTCCCAACCTCTCCCAATGCTCCCT-----GACACGATACCTTTACGACCTG 1824
Qy 565 AspAspIleGlyThrIleSerGlyTrpGlyLeuThrGlnArgGlyPheLeuAlaArgAsn 584
Db 1825 GCGTTATGAGCATGTGATGCGCTTCCGGGTTATGAGAGAG---ATTGCTCATGAC 1881
Qy 585 LeuMetTrpValAspIleProIleValAspHisGlnLysCysThrAlaIleArgLys 604
Db 1882 CTCAGCTTTGTCGTCGCTCCCGTACGTAATCCACAGCGCTGT----- 1923
Qy 605 ProProTrpProArgGlySer-----ValThrIleAsnMetLeuCysAla 619
Db 1924 GAGAACTGGCTCCGGGGAAGAATAGCATGATGTCTTCCTCAAAACATGTTCTGTGCT 1983
Qy 620 GlyLeuGlnSerGlyLysAspSerCysArgGlyAspSerGlyGlyAlaLeuValPhe 639
Db 1984 GGCACACCATCTCTTAAGCAGACGCCCTGCCAGGGGATGTTGGGCGCTTTTGCAGTA 2043
Qy 640 LeuAspSerGlyThrGlnArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsn 659
Db 044 AGCGACCCCAACACATGATGCTGGTGGCCACGCGCATGCTGCTCGGGGC---ATCGGG 2100
Qy 660 CysGlyGlnAlaGlyLeuTrpGlyValIleTrpThrLysValIleAsnTrpIleProTrpIle 679
Db 2101 TGCAGCAGGGGC-----TATGCTTCTACACCAAGTGTCTCAACTGACTGACTGATC 2154
Qy 680 Gln 680
Db 2155 AAG 2157

RESULT 2
US-09-027-337-1
; Sequence 1, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirofumi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; NATURE:
; APPLICATION: 23,2589
; OTHER INFORMATION: cDNA sequence of TADG-15
US-09-027-337-1

Alignment Scores:
Pred. No.: 2 89e-40 Length: 3147
Score: 490.00 Matches: 186
Percent Similarity: 37.81% Conservative: 90
Best Local Similarity: 25.48% Mismatches: 270
Query Match: 13.04% Indels: 184
DB: 2 Gaps: 31

US-09-874-198-2 (1-686) x US-09-027-337-1 (1-3147)
Qy 29 ArgLeuAlaSerProGlyPhePro---GlyGlyTrpAlaAsnAspGlnIleuArgTrp 47
Db 701 CGCTTCACACAGCGCGCTTCCCTGACAGCCCTTACCCCGCTATGCCCCCTCCAGTGG 760
Qy 48 ThrLeuThrAlaProProGlyTrpArgLeuArgLeuTrpPheThrHisPheAspLeuGlu 67
Db 761 GCGCTCGGGGGAGACCCGACATCAGTGTGAGCCTCACCTCCGACGCTTTGACCTTGGC 820
Qy 68 LeuSerHisLeuGlySerGlyTrpPheValLysLeuSerSerGlyAlaLys----- 84
Db 821 TCTGTGGACAGCGCGGACAGCGACCTGTGTACGCTGTAACAACACCTGAGCCCATGGAG 880
```

```
Qy 85 -----ValLeuAlaThrLeuGlyGlnGlnIleSerThrAspThrGlnArgAlaProGly 102
Db 881 CCCCAGCCCGCTGTGAGATGTTGTGGC-----ACCTAACCTCCCTCCCTAG 925
Qy 103 LysAspThrPheTrpSerLeuGlySerSerLeuAspIleThrPheArgSerAspTrpSer 122
Db 926 AACCTGACCTTCCACATCCGCCAGACGTCCTCTGATACATGATTAACCAACACTGAG 985
Qy 123 AsnGlnLysProPheThrGlyPheGlnAla---PheTrpAlaIleAlaLysPheArgGln 141
Db 986 CGCGGCGCATGCC-----GGCTTTGAGCCACCTCTTCCAGCTGCTAGAGATGACACC 1039
Qy 142 CysGlnValAlaProGlyGlnAlaProThrCysAspHisHisCysHisAsnHisLeuGly 161
Db 1040 TGT----- 1042
Qy 162 GlyPheTrpCysSerCysArgAlaGlyTrpValLeuHisArgAsnLysArgTrpCysSer 181
Db 1042 ----- 1042
Qy 182 AlaLeuGlySerGlyGlnValPheThrGlnArgSerGlyLysLeuSerSerProGlyTrp 201
Db 1043 -----GGAGCCCGCTTACGTAAAGCCAGGGGACATTTCAACAGCCCTACTAC 1090
Qy 202 ProArgProTrpProLysLeuSerSerCysThrTrpSerIleSerLeuGlnGlnGlyPhe 221
Db 1091 CGAGGCCACATGACCCCAACATGATGACTGCACATGGAACATTTAGAGTGGCCCAACACAG 1150
Qy 222 SerValIleLeuAspPheValGlnSerPheAspValGlnThrHis---ProGlyThrLeu 240
Db 1151 CATGTGAAGGTGAGCTTCAAAATCTTCTACCTGTGAGACCGCCGCTGCTGGCGGACACC 1210
Qy 241 CysProTrpArgPheLeuLysIleGlnThrAspArgGlnLysIleGlyProPheCysGly 260
Db 1211 TGCCTCAAGACTACGTGGAGATCAATGGGAGAA-----TACTCGGA 1255
Qy 261 LysThrLeuProHisArgIleGlyThrLysSerAsnThrValThrIleThrPheValThr 280
Db 1256 GAGAGGTCCACATGCTGCTGCACACACACAGACACAGACATGATCACTGCTTCCACTCA 1315
Qy 281 AspGlnSerGlyAspHisThrGlyTrpLysIleHisTrpThrSer-----ThrAlaGln 298
Db 1316 GATCAGTCTACACACGACACCGGCTTCTTAACTGATACCTCTCTCAACGACTCCAGTAC 1375
Qy 299 ProCysProTrpProMetAlaProProAsnGlnHisValSerProValGlnAlaLysTrp 318
Db 1376 CCATGCCCG-----GGGCACTTCACTGCGCCGACGCGCGGTGT 1414
Qy 319 IleLeuLys-----AspSerPhe 324
Db 1415 ATCCGGAAGAGCTGCGCTGTGATGGCTGGCGGACATGCACACCAACAGCATGAGCTC 1474
Qy 325 SerIlePheCysGlnThrGlyTrpGlnLeuLeuGlnGlnHisLeuProLeuLysSerPhe 344
Db 1475 AACCTGAGTTGCGACCGCGCCACAGTTACAGTCAAGAAAGTTTTCGCAAGCCCTTC 1534
Qy 345 ThrAlaValCysGlnLys-----AspGlySerTrpAspArgProMetPro 359
Db 1535 TTCTGGGTCTGCGACAGTGTGAACGACTGCGGAGACAAACAGCGACGAGC----- 1585
Qy 360 AlaCysSerIleVal----- 364
Db 1586 GGGTGCAGTTGTGCGGCGCCAGACCTTCAAGTGTCCAAATGGAAGTCCCTCGAANAAGC 1645
Qy 365 -----AspCysGlyPro---ProAspAspLeuProSerGlyArg 376
Db 1646 CAGCACTGCAATGGAAGACAGCACTGTGGGAGCGGTCCGACAGAGCGCTTCCTCCCAAG 1705
Qy 377 ValGlnTrpIlePheGlyProGlyValIleThrThrLysAlaValIleGlnTrpCys 396
Db 1706 GTGAAGCTGTACTTGTACAAACACACC-----TACCGCTGC 1744
```

```

QY 397 GUGLUthRhetYrThrmElysValasn-----AspElyLysTyRValCysGLu 413
      ||| ||| ||| ||| ||| |||
Db 1745 CPGAAAGGGCTCGCTTGACCAAGGCAACCCGAGTGTACGGGAAGGAGCTGT--- 1801
QY 414 AlaAspGlyPheTrpThrSerSerLysGlyLysSerLeuProValCysGluProVal 433
      :||:||||| :||:|||||
Db 1802 AGCGACGGC-----TCAGATGAGAGAGAC-----TGGCAG----- 1831
QY 434 CysGlyLeuSerAlaArgTrpThrGlyLysArgIleTyRLeuGlyLysGlnLysAlaLysPro 453
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1832 TGTGGCTGGGGTCATTCACAGACAGACGGCTCGTGTGGGGGCACGGATGGATGAG 1891
QY 454 GlyAspPheProTrpGlnValLeuLeu-----LeucGly---GlyThrAlaAlaGly 470
      |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 1892 GCGGAGTGGCCCTGGCAGGTAAGCTGCATGCTGTGGGCCACAGGCCACATCTCGGTTGCT 1951
QY 471 AlaLeuLeuTyRAspAsnTrpValLeuThrAlaAlaHisAlaValGlyGlnGlnLys--- 489
      :||:||||| :||:||||| :||:||||| :||:||||| :||:||||| :||:|||||
Db 1952 TCCCTCATCTCTCCCACTGGCTGTCTGTCCGACACTGTACATCGATGACAGAGA 2011
QY 490 -----HisAsp----- 491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2012 TTACAGTACTCACACCCACCGACGTGACGGCTTCTGGCTGGCTTGCACGACAGACCCAG 2071
QY 492 AlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTyRThrGln 511
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2072 CGCAGGCGCCCTGGGGTGAGGAGCGCAGGCTCAAGGCGCATCTGCCAC----- 2122
QY 512 AlaTrpSerGluAlaValPheIleHisGluGlyTyRThrHisAspAlaGlyPheAspAsn 531
      :||:||||| :||:||||| :||:||||| :||:||||| :||:||||| :||:|||||
Db 2123 -----CCCTTCTTCATGATGACTTTCACCTTCGACTAT 2152
QY 532 AspIleAlaLeuIleLysLeuAsnLysValValIleAsnSerAsnIleThrProIle 551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2153 GACATGGCGCTGCTGGAGCTGGAGAAACCGGACAGACTACAGCTCCATGTGTGGCCCATC 2212
QY 552 CysLeuProArgLysGluAlaGlnUserPheMetArgThrAspAspIleGlyThrAlaSer 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2213 TGGCTTCGCG-----GACGCTCCCATGTCTCCCTGCGCGCAAGGCCATCTGGTCCAG 2266
QY 572 GlyTrpGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetGlyValAspIle 590
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2267 GGGTGGGGACACCCAGATGAGGACACTGGCGCGCTGATCTCTGCAAAAGGTGAGATC 2326
QY 591 ProIleValAspHisGlnLysCysThrAlaAlaTyRGlutLysProProtyrProArgGly 610
      :||:||||| :||:||||| :||:||||| :||:||||| :||:||||| :||:|||||
Db 2327 CGGGTCAACACAGACCACTCGAGAACCTCTGCCGAG----- 2368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 611 SerValThrAlaAsnMetLeuCysAlaGlyLeuGlnSerGlyLysAspSerCysArg 630
      :||:||||| :||:||||| :||:||||| :||:||||| :||:||||| :||:|||||
Db 369 CAGATGACCGCGCATGATGTGCTGGCTTCCCTGACGCGCGGTGAGCTCTGCTCCAG 2428
QY 631 GlyAspSerGlyLysAlaLeuValPheLeuAspSerGluThrGlnArgTrpPheValGly 650
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2429 GGGATTCGCGGGGACCCCTGTCCACGCTGAGGCGCAT---GGGCGGATCTTCCAGGCC 2485
QY 651 GlyIleValSerTrpLysSerMetAsnCysGlyLysAlaGlyLysGlnTyRThrGly 670
      |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 2486 GGGTGTGTAGCTGGGAGAGC---GGCTGGCTCAGAGAGAACAGCAGGCGTGTACACA 2542
QY 671 LysValIleAsnTyRLeuProTrpIleGlu 680
      :||:||||| :||:||||| :||:||||| :||:||||| :||:||||| :||:|||||
Db 2543 AGGCTCCCTCTGTTCGGGACTGTGATCAAA 2572

```

```

; FILE REFERENCE: D6064CTP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-1

Alignment Scores:
Pred. No.: 2,89e-40 Length: 3147
Score: 490.00 Matches: 186
Percent Similarity: 37.81% Conservative: 90
Best Local Similarity: 25.48% Mismatches: 270
Query Match: 13.04% Indels: 184
DB: 4 Gaps: 31

US-09-874-198-2 (1-686) x US-09-644-600-1 (1-3147)
QY 29 ArgLeuAlaSerProGlyPhePro---GlyGluTyRAlaAsnAspGlnGlnArgTrp 47
      ||| :||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 CGCTTACACAGCGCGGCTTCCCTGACAGCCCTTACATCCCGCTCATCCCGCTGCGCAGTGG 760
QY 48 ThrLeuThrAlaProProGlyTyRArgLeuArgLeuTyRThrHisPheAspLeuGlu 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 761 GCCCTGGGGGGAGCCCGCATGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 820
QY 68 LeuSerHisLeuCysGluTyRAspPheValLysLeuSerSerGlyAlaLys----- 84
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 821 TCCCTGGAGAGCGCGGACGCGCTGTGACGCTGTACACACCCCTGAGCCCATGAG 880
QY 85 -----ValLeuAlaThrLeuCysGlyGlnGlnUserThrAspTrpGlnArgAlaProGly 102
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 881 CCCACGCCCTGGTGGAGTGTGTGTC-----ACCTAACCTCTCTCTTAC 925
QY 103 LysAspThrPheTyRSerLeuGlySerSerLeuAspIleThrPheArgSerArgTrpSer 122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 926 AACCTGACCTTCCACATCTCCCGACAGACGTCGTCGTCATGATACCAACTGAG 985
QY 123 AsnGlnLysProPheThrGlyPheGluAla---PheTyRAlaAlaGlnAspIleAspGlu 141
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 986 CGCGGCGATCCC-----GGCTTGGAGGCCACCTTCTTCCAGCTGCGCTGAGTACAGCAGC 1039
QY 142 CysGlnValAlaProGlyGlnAlaProThrCysAspHisHisCysHisAsnHisLeuGly 161
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1040 TGT----- 1042
QY 162 GlyPheTyRAspSerCysArgAlaGlyTyRValLeuHisArgAsnLysArgTrpCysSer 181
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 ----- 1042
QY 182 AlaLeuCysSerGlyLysValPheThrGlnArgSerGlyLysLeuSerSerProGlyTyR 201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1043 -----GAGGCGCGCTTACGTAAGCCACAGGAGACATTCACAGCCCTTACTAC 1090
QY 202 ProArgProTyRProLysLeuSerSerCysTrpTrpSerIleSerLeuGlnGluGlyPhe 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1091 CCAAGCCACTACCCACCAACATTCAGTACGTACATGAGACATGAGAGTGGCCCAACAACAG 1150
QY 222 SerValIleLeuAspPheValGlnUserPheAspValGluThrHis---ProGluThrLeu 240
      ||| :||:||||| ||| :||:||||| ||| :||:||||| ||| :||:||||| ||| :||:|||||
Db 1151 CATGTGAGGTAGAGCTTCAAAATTCTTCTACCTGTGACCCCGGCGCTGCGGCGAC 1210
QY 241 CysProTyRAspPheLeuLysIleGlnTrpAspArgGlnGlnHisGlyProPheCysGly 260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1211 TGGCCCAAGGACTACGTGAGATCATGTGGGAGAA-----TACTGCCGA 1255

```

```

RESULT 3
US-09-644-600-1
; Sequence 1, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; OVEREXPRESSED IN CARCINOMAS

```

```

QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr 280
   :::          :::          |||||          ::|||::          |||
Db 1256 GAGAGTCCACAGTGTCTGTACACACAGACAGACAAATACACTGTGCTTCCACTCA 1315
QY 281 AspGluSerGlyAspHisThrGlyTrpLysIleHisThrSer-----ThrAlaGln 298
   |||||::          |||||::          |||
Db 1316 GATCACTCTTACACACACACCGCTTCTTACGTAGTACTCTCTTACAGACTCCACTGAC 1375
QY 299 ProGlyProThrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysThr 318
   |||||::          |||
Db 1376 CCATGCCC-----GGGCACTTACAGTCCGCGCACAGGGCGGTGT 1414
QY 319 IleLeuLys-----:::-----:::-----AspSerPhe 324
   |||
Db 1415 ATCCGGAAGAGAGTGGCTGTATGTGTGGTGGCCGACTGCACCGACACAGCGATGAGCTC 1474
QY 325 SerIlePheCysGluThrGlyTrpGluLeuLeuGlnGlyHisLeuProLeuLysSerPhe 344
   :::          |||::          |||
Db 475 AACTGCAGTTGGACGCGCGCCGACACAGTTCCAGTCAGACAAAGTTTCTGCAAGCCCTC 1534
QY 345 ThrAlaValCysGlnLys-----AspGlySerThrPaspArgProMetPro 359
   |||||
Db 1535 TTCTGGGTCTGCACAGTGTGAACACACTGCGGAGACACAGCAGCAGAG----- 1585
QY 360 AlaCysSerIleVal-----:::-----364
   |||||
Db 1586 GGGTCGAGTTGTCCGGCCAGACCTTCAGGTGTTCATGGGAGAGTCCCTTCGAAAGC 1645
QY 365 -----AspCysGlyPro--ProAspSpleupProSerGlyArg 376
   |||||
Db 1646 CAGCACTGCATATGGAGAGACGACCTGTGGGAGCGGGTCCGACAGCGCTCTCCCAAG 1705
QY 377 ValGluTrpIleThrGlyProGlyValThrThrTrpLysAlaValIleGlnTrpSerCys 396
   |||
Db 1706 GTGACGCTGTCACTTGTACCAACACACC-----TACCCTGC 1744
QY 397 GluGluThrPheThrMetLysValAsn-----AspGlyLysTrpValCysGlu 413
   |||
Db 1745 CTCATGGGCTGCTGTGAGCAAGGCAACCTGACTGTGACGGAGGAGGAGCTGT--- 1801
QY 414 AlaAspGlyPheThrPheSerSerLysGlyGluLysSerLeuProValCysGluProVal 433
   ::|||
Db 1802 AGCGAGGG-----TCAGATAGAAAGAC-----TCCGAC----- 1831
QY 434 CysGlyLeuSerAlaArgThrThrGlyArgIleTrpGlyGlyGlnLysAlaLysPro 453
   |||||
Db 832 TGTGGCTCGGCTCATTCACAGACAGGCTCGTGTGTGGGGGACAGGATGCGGATGAG 1891
QY 454 GlyAspPheProTrpGlnValLeuIle-----LeuGly---GlyThrThrAlaAlaGly 470
   |||||
Db 1892 GGGGACTGGCCCTGCGAGTAAGCTTCATGCTGCGGCGACGCGACATCTCGGTGCT 1951
QY 471 AlaLeuLeuTrpAspAsnTrpValLeuThrAlaAlaHisAlaValTrpGluGlnLys--- 489
   ::|||
Db 1952 TCCCTCATCTCTCCCAACGTGGGTCTCTGCGCACACTGTCTACATCCATGACAGAGA 2011
QY 490 -----HisAsp-----491
   |||||
Db 2012 TTCAGTACTCAGACCCACGACGAGGAGCGCTTCTCGGCTTGCAGCGACAGCCAG 2071
QY 492 AlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTrpThrGln 511
   |||||
Db 2072 CCCAGCGCCCTGGGTGGAGGAGCGCAGGCTCAAGCGATCATCTCCACC----- 2122
QY 512 AlaTrpSerGluAlaValPheIleHisGluGlyTrpThrHisAspAlaGlyPheAspAsn 531
   ::|||
Db 2123 -----CCCTTCTTCAATGACTTTCACCTTCGACTAT 2152
QY 532 AspIleAlaLeuIleLysLeuAsnAsnLysValValIleAsnSerAsnIleThrProIle 551
   |||||
Db 2153 GACATGGCGCTGTGAGCTGGAGAAACCGGCGAGAGTACAGCTTCATGTGTGCGCCCATC 2212

```

```

QY 552 CysLeuProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSer 571
   |||||
Db 2213 TGGCTCCG-----GACGCTCCCATGTCTTCCCTCGCGGGAAGCCATCTGGGTGACG 2266
QY 572 GlyTrpGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetLysValAspIle 590
   |||||
Db 2267 GGTGGGACACACCCAGATGAGGACACTGTGCGGCTGATCTTCCAAAGGCTGAGATC 2326
QY 591 ProIleValAspIleGlnLysCysThrAlaAlaValGlyGluLysProProArgGly 610
   ::|||
Db 2327 CCGGTATACACACAGACCACTTCCGAGAACCTCTCCGCGAG----- 2368
QY 611 SerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyLysAspSerCysArg 630
   ::|||
Db 2369 CAGATACAGCGCGCATGATGTCTCGTGGCTTCTCAGCGCGCGGTGACCTCTCCAG 2428
QY 631 GlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThrGluArgTrpPheValGly 650
   |||||
Db 2429 GGTGATTCGCGGGGACCCCTGTCCACCGTGAAGCGGAT---GGCGGATCTTCCAGGCC 2485
QY 651 GlyIleValSerTrpGlySerMetAsnCysGlyLysAlaGlyGlnTrpGlyValTrpThr 670
   |||||
Db 2486 GGTGTGTGAGCTGGGGAGAC---GGCTGCGCTCAGAGACAAACGACGAGCGTGTACCA 2542
QY 671 LysValIleAsnTrpIleProTrpIleGlu 680
   ::|||
Db 2543 AGGCTCCCTCTGTTCGGGAGCTGATCAAA 2572

RESULT 4
US-09-644-600-18/c
; Sequence 18, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsu
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 18
; LENGTH: 3147
; TYPE: RNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Antisense of TADG-15
US-09-644-600-18

Alignment Scores:
Pred. No.: 2,89e-40 Length: 3147
Score: 490.00 Matches: 186
Percent Similarity: 37.81% Conservative: 90
Best Local Similarity: 25.48% Mismatches: 270
Query Match: 13.04% Indels: 184
DB: 4 Gaps: 31

US-09-874-198-2 (1-686) x US-09-644-600-18 (1-3147)
QY 29 ArgLeuAlaSerProGlyPhePro---GlyGluTrpAlaAsnAspGlnGluArgArgTrp 47
   |||
Db 2447 CGCTTACACAGCGCGCTTCTGTACAGCCCTTACCCCTCATATCCCGCTTCAGTGG 2388
QY 48 ThrLeuThrAlaProProGlyTrpArgLeuArgLeuTrpPheThrHisPheAspLeuGlu 67
   |||
Db 2387 GCCCTCGGGGGAGCCCGACATAGTGTGAGCTTACCTTCCGCAAGCTTGTGACTTGGC 2328
QY 68 LeuSerHisLeuCysGluTrpAspPheValLysLeuSerSerGlyAlaLys----- 84
   |||

```

Db 2327 TCCTGGACAGCGCGACGACGACGTCGTGACGCTGACACACGACCTGAGCCCATGAG 2268
 QY 85 -----ValleuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAlaProGly 102
 Db 2267 CCCACGCGCTGGTGCATGTTGTGGC-----ACCTAACCTCCCTCTCTAC 2223
 QY 103 LysAspThrPheThrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrSer 122
 Db 2222 AACCTGACCTTCACATGCTCCCGACAGACGTCGTCATCATCACTGATACACACACTGAG 2163
 QY 123 AsnGluLysProPheThrGlyPheGluAla---PheTyrAlaIleGluAspIleAspGlu 141
 Db 2162 CGCGCGCATCC-----GGCTTTGAGCGCACCTCTCTGACGCTGCTAGAGACAGC 2109
 QY 142 CysGlnValAlaProGlyAlaIleProThrCysAspHisHisCysHisAsnHisLeuGly 161
 Db 2108 TGT----- 2106
 QY 162 GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSer 181
 Db 106 ----- 2106
 Db 182 AlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGlyTyr 201
 Db 2105 -----GGAGCGCGCTTACGTAAGCCGAGGCGGACATTCACACGCCCTACTAC 2058
 QY 202 ProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGlnGluGlyPhe 221
 Db 2057 CCAGGCCACTACCCACCCACACATTGACTGACATGGACATTTGAGGTGCCCCAACACACAG 1998
 QY 222 SerValIleLeuAspPheValGluSerPheAspValGluThrHis---ProGluThrLeu 240
 Db 1997 CATGTGAGGTGAGTCTCAATCTCTTACTACCTGTGAGCCGCGGCTGGGGGAC 1938
 QY 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGly 260
 Db 1937 TGCCCAAGACACTGATGGAGATCAATGGGAGAAA-----TACTCGCGA 1893
 QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValIleThrPheValThr 280
 Db 1892 GAGAGGTCCAGTTCGTGCTGCACACACACACACACACAGTTCGCTTCCACTCA 1833
 QY 281 AspGluSerGlyAspHisThrGlyTyrLysIleHisTyrThrSer-----ThrAlaGln 298
 Db 1832 GATCAGTCTCCACACGACACGCGCTCTTACGTAATCACTCTCCACGACTCCAGTAC 1773
 QY 299 ProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyr 318
 Db 1772 CCATGCCG-----GGGAGTTCACGTGCCGCGAGGGGGGTGT 1734
 Db 319 IleLeuLys-----AspSerPhe 324
 Db 1733 ATCCGGAAGAGTGGCGTGTGATGGCTGGCGGACTGCACGACACACAGCGATGAGCTC 1674
 QY 325 SerIlePheCysGlyThrGlyTyrGluLeuLeuGlnGlyHisLeuProLeuLysSerPhe 344
 Db 1673 AACGTGAGTTGCGACGCGCGCACAGTTCACGTGCAAGAACAAAGTTCTGCAAGCCCCC 1614
 QY 345 ThrAlaValCysGlnLys-----AspGlySerThrPaspArgProMetPro 359
 Db 1613 TTCTGGGTCTGCACAGTGTGAACGACTGGGAGACACAGCGAGGAG----- 1563
 QY 360 AlaCysSerIleVal----- 364
 Db 1562 GGGTGCAGTTGTCCGGCCGACACTTCAGGTTCCTCAATGGAGAGTCCCTCTGAAAGC 1503
 QY 365 -----AspCysGlyPro---ProAspAspLeuProSerGlyArg 376
 Db 1502 CAGCAGTGAATGGAAGAGACGACTGTGGGAGCGGTCCGACGAGCGCTCTCCCCAG 1443
 QY 377 ValGluTyrIleThrGlyProGlyValIleThrTyrLysAlaValIleGlnTyrSerCys 396
 Db 1442 GTGAACGTGCTACTTGTACCAAAACACACC-----TACCGCTGC 1404

QY 397 GluGluThrPheThrThrMetLysValAsn-----AspGlyLysTyrValCysGlu 413
 Db 1403 CTCATGGGCTGCTGCTTGAACAGGCGACACCTGAGTGTGACGAGGAGGAGACTGT--- 1347
 QY 414 AlaAspGlyPheThrThrSerSerLysGlyGlyLysSerLeuProValCysGluProVal 433
 Db 1346 ACCGACGCG-----TCAGATGAGAAAGAG-----TGCAGC----- 1317
 QY 434 CysGlyLeuSerAlaArgThrThrGlyArgIleTyrGlyGlyGlnLysAlaLysPro 453
 Db 1316 TGTGGGTGCGGTCACTACAGACAGAGCTGTGTGTGGGGGACAGGATCGGATGAG 1257
 QY 454 GlyAspPheProThrGlnValLeuIle-----LeuGly---GlyThrThrAlaIleGly 470
 Db 1256 GCGAGTGGCGCTCGGAGGTAAACCTGCATGCTGTGGCGGACAGGCGCACATTCGGGTGT 1197
 QY 471 AlaLeuLeuTyrAspAsnThrValLeuThrAlaAlaHisAlaValTyrGluGlnLys--- 489
 Db 1196 TCCCATCTCTCCCAACTGGCTGGTGTCTGGCGGACACTGCTACATGATGACAGAGGA 1137
 QY 490 -----HisAsp----- 491
 Db 1136 TTCAGTACTACAGCCGACGAGTGGAGCGCTTCTGCGCTTGACACACAGAGCCAG 1077
 QY 492 AlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGln 511
 Db 1076 CGCAGCGCCCTGGGTGGAGAGCGGACGCTCAGAGCATATCTCCAC----- 1026
 QY 512 AlaThrSerGluAlaValPheIleHisGluGlyTyrThrHisAspAlaGluPheAspAsn 531
 Db 1025 -----CCCTCTCAATGACTTCACCTTCGACTAT 996
 QY 532 AspIleAlaLeuIleLysLeuAsnAsnLysValValIleAsnSerAsnIleThrProIle 551
 Db 995 GACATCGCGCTGTGGAGTGAAGAAACCGGACAGTACAGTCAATGATGGCGCCATC 936
 QY 552 CysLeuProArgGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSer 571
 Db 935 TGCTGTCCG-----GAGCGCTCCCATGTCTTCCTGCGCGGAGAACGCTTCGGGTACG 882
 QY 572 GlyThrGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetTyrValAspIle 590
 Db 881 GGTGTGGGACACACCAAGTATGGAGGACACTGGCGGCTGATCTGCAGAAAGGCTGAGATC 822
 QY 591 ProIleValAspHisGlnLysCysThrAlaIleTyrGluLysProProTyrProArgGly 610
 Db 821 CGGTCATCAACGACACACCTGCGAGAACTCTGCGGAG----- 780
 QY 611 SerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGlyLysAspSerCysArg 630
 Db 779 CAGATACGCGCGCATGATGTGCGTGGCTTCTCAGGCGGCGGTGACCTCTGCCAG 720
 QY 631 GlyAspSerGlyGlyAlaLeuValPheLeuAspSerGlyThrGluArgThrPheValGly 650
 Db 719 GGGATTTCCGGGGGACCCCTGTCACGCTGGAGGGGGAT---GGCGGATCTTCCAGGCC 663
 QY 651 GlyIleValSerThrGlySerMetAsnGlyGluAlaGlyGlnTyrGlyValTyrThr 670
 Db 662 GGTGTGTGAGCTGGGGAGAC---GGCTGCTCTCAGAGAACACGACGAGCTGTACACA 606
 QY 671 LysValIleAsnTyrIleProThrIleGlu 680
 Db 605 AGGCTCCCTCTGTTCGGGACTGATCAAA 576

RESULT 5
 US-08-296-014A-3
 ; Sequence 3, Application US/08296014A
 ; Patent No. 5716834
 ; GENERAL INFORMATION:
 ; APPLICANT: Ding, Jeek Ling
 ; APPLICANT: Ho, Bow
 ; TITLE OF INVENTION: The Cloned Factor C cDNA of the


```

Db 2838 TTACCACTG--ACAGTAACAGAGAACATGCTGTGCAGCTTACAGAAAGGACCTTAT 2894
QY 627 AspSerCysArgGlyaspSerGlyGlyAlaLeuValPheLeu--AspSerGluThrGlu 645
Db 2895 GATGCCCTGAGTGTGGAGAGGATGTGCAGTGGGCGAGTCCCGATGATCCCTACCGAA 2954
QY 646 --ArgTrpPheValGlyGlyIleValSerTrpGlySer--MetAsnCysGlyGluAla 663
Db 2955 AGCGGGGTGGTCTTGGAGAGGATGTGCAGTGGGCGAGTCCCGATGATGTCGCAAGCG 3014
QY 664 GlyGlyTrpGlyValGlyThrLysValIleAsnTrpIleProTrpIleGluAsnIleIle 663
Db 3015 AACCACTACGGGGCTTCACTAAAGTTAACGTTTCTGTCATGATTTAGGACATTCAAT 3074

RESULT 6
US-08-596-405-3
; Sequence 3, Application US/08596405
; Patent No. 5858706
; GENERAL INFORMATION:
; APPLICANT: Ding, Jeak Ling
; TITLE OF INVENTION: The Cloned Factor C cDNA of the
; TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscorpius
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,405
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1781-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3448 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Carcinoscorpius rotundicauda
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 18..3074
; US-08-596-405-3

Alignment Scores:
Pred. No.: 3,46e-38 Length: 3448
Score: 470.50 Matches: 151
Percent Similarity: 39.63% Conservative: 63
Best Local Similarity: 27.96% Mismatches: 185
Query Match: 12.52% Indels: 141
DB: 2 Gaps: 26

```

```

US-09-874-198-2 (1-686) x US-08-596-405-3 (1-3448)
QY 265 HisArgIleGluThr-----LysSerAsnThrVal 274
Db 1515 CACAGACTGAGGCTGAGAGAGCCCTTTATTTGGAGTAATGATGATGTATGTGTT 1574
QY 275 -----ThrIleThrPhe--ValThrAspGluSerGlyAspHisThrGly----- 288
Db 1575 CTGAATGATTAACCTAATCATTTCTGGGCTGTGGCAACCTGGAATAATGAACAGTGTA 1634
QY 289 -----TrpLysIleHisTrpSerThrAla 297
Db 1635 TATATGACATCCAGATCATGTTCAGTCTGTGTGGAATAAC-----AAGTCAGTGTTT 1688
QY 298 GlnPro-----CysPro 301
Db 1689 CAGCCCTCAAGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1748
QY 302 TyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTrpIleLeuLys 321
Db 1749 GATCCTGATCAGTGGAAATGAGACGCCACA-----CTTCATGAGCAAAATGTT----- 1799
QY 322 AspSerPheSerIlePheCysGluThrGlyTyrglyLeuLeuGlnGlyHisLeuProLeu 341
Db 1800 GATGGGTCTATGCTGTGTTCTTATATAGTACAGCTGTGAGGTTCTCCACTACCTCAGT 1859
QY 342 LysSerPheThrAlaValGlyGlnLysAspGlySerTrpAspArgProMetProAlaCys 361
Db 1860 GGAAGTGAACCGTAACCTGTCACACAAATGAGAGTGGAGTCTCTTAACCTGCATGT 1919
QY 362 --SerIleValAspCys--GlyProProAspLeuProSer--GlyArgValGln 378
Db 1920 ATCAAGTATCATCCCTGCCCAAAACCCCT-----GTACCATCATATGTTCTGTGGAA 1973
QY 379 TyrIleThrGlyProGlyValThr----- 387
Db 1974 --ATCAAAACCCCAAGTGGAGCAAACTGATAGTGTGTGGTGCACCTTTCTTGAGG 2030
QY 387 ----- 387
Db 2031 TTGCCAGGTTACCCCTCCCATTTAGCTAGACGACCAACCTCTCCAAACCTAGATCC 2090
QY 387 ----- 387
Db 2091 TCACAAACCTCACTAGTGGACTTGGCTTCTAAAGTTAACTACCTGAAGCTATACGG 2150
QY 388 -----TyrLysAlaValIleGlnTrpSerCysGluGluThrPheTyrThrMetLysVal 405
Db 2151 GTAGGCTCTCGAGCATC-----TACAGCTGCGAGTCCAGATACTACCACTACTTGA 2204
QY 406 AsnAspGlyLysTrpValGlyCysGlnAlaAspGlyPheTrpThrSerSerLysGlyLys 425
Db 2205 TCTCAAGGCAGAGA--TGTGACTATATGAAAGTGAAGTGGTGGCCAGCAGC-- 2258
QY 426 SerLeuProValCysGluProValCysGlyLeuSerAlaArgThrThrGlyLysArgLe 445
Db 2259 -----TGTATTCAGTTTGTGGACGGTGAAGTCTCTCTCTCTCTCTCTCTTATC 2306
QY 446 TyrGlyGlyGlnLysAlaLysProGlyAspPheProTrpGlnValLeuIle----- 462
Db 2307 TCGAATGGAATTTACAAATAATAGTCAAGTCCGTCGAGCAGAGAAATCTCTAGATGG 2366
QY 463 -----LeuGlyGlyThrThrAlaIleAlaIleLeuLeuLysPaspAsnTrp 477
Db 2367 CTTCGAGACCAATATATGTTGTTTCTCCAGTGTGGAGATCTCTATGATGAGAAATGG 2426
QY 478 ValLeuThrAlaAlaHisAlaValTyrglyGlnLysHisAspAlaSerAlaLeuAspIle 497
Db 2427 ATGTCATGCTGCGCCACGTGTCT-----ACCTATCTGCTACGTCTGAGATTAT 2477
QY 498 ArgMetGlyThrLeuLysArgLeuSerProHisTrpThrGlnAlaTrpSer----- 514

```


Db 2478 GACCCCAATCAGTTTAAATGATCTGCGCACTACCGTGATGACATAGACGAT 2537
OY 515 -----GluAlaValPheIleHisGluClyTyrThrHisAspAlaGly 528
Db 2538 GACTATGTACAGTAAGAGAGGCTCTTGAGATCAGTGAATCTCACTACGACCCGGC 2597
OY 529 PheAspAsn---AspIleAlaLeuIleLysLeuAsnAsnValValIleAsnSerAsn 547
Db 2598 AATCTCACTTTGACATAGACCTTAATTCAAACTGAAACCTCTTACTTGTGACACACGA 2657
OY 548 IleThrProIleGlyLeuProAlaGly---GluAlaGlySerPheMetAlaArgAsnLeuMet 566
Db 2658 GTCCAAACCAATCTGTCTGCTACTGATCACAACAAGAACACCTATTTCAGAGACGATTCAA 2717
OY 567 IleGlyThrAlaSerGlyTyrPglLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMet 586
Db 2718 TTAGCAGTGTGACAGTTTGGGTTTGATGAATAACAACACCTATTTCAGAGACGATTCAA 2777
OY 587 TyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProPro 606
Db 778 CAAGCTGTCTACCTGTGTTGTCAGCAGCAGCAGCTGTGAAGAGGGGTACAAAGAACGAC 2837
OY 607 TyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGlySerGlyGlyLys 626
Db 2838 TTACCACTG---ACAGTACACAGAAACATGTTCTGTGACAGTTTACAAAGAGGACGTTAT 2894
OY 627 AspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeu---AspSerGluThrGlu 645
Db 2895 GATGCTGTGACGTGGGACAGTGAAGACCTTTAGTGTGTGCTATGATGCTCCCTACCGAA 2954
OY 646 ---ArgTyrPheValIleGlyIleValSerTyrPglSer---MetAsnCysGlyGluAla 663
Db 2955 AGCGGTGGTGTGGAGAGGATGTGACGTGGGACAGTCCAGTGGATGTGGCAAGCGC 3014
OY 664 GlyGlnTyrGlyAlaTyrThrLysValIleAsnTyrIleProThrIleGluAsnIleIle 683
Db 3015 AACCACTACGGGGCTTCACTAAAGTTAACGTTTCTGTCTGATGATTTAGGACGTTCAAT 3074

RESULT 7
US-08-877-620-3
Sequence 3, Application US/08877620
Patent No. 5985590
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
APPLICANT: HO, Bow
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinus scorpis
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877,620
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/596,405
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ. ID NO. 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3448 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Carcinus scorpis rotundicauda
FEATURE:
NAME/KEY: CDS
LOCATION: 18..3074
US-08-877-620-3

Alignment Scores:
Pred. No.: 3,466-38
Score: 470.50
Percent Similarity: 39.63%
Best Local Similarity: 27.96%
Query Match: 12.52%
DB: 2
Gaps: 26

US-09-874-198-2 (1-686) x US-08-877-620-3 (1-3448)

OY 265 HisArgIleGluThr-----LysSerAsnThrVal 274
Db 1515 CACAGACTAAGCTCTGAGAACCCCTTATTGGAGTTAATGATGATGATGATGATGAT 1574
OY 275 -----ThrIleThrPhe---ValThrAspGluSerGlyAspHisThrGly----- 288
Db 1575 CTGAATGATTAACCTAACATCTGCGCTGTGGCAACCTGGAATGAATGAATGAATGAAT 1634
OY 289 -----TyrPheIleHisTyrThrSerThrAla 297
Db 1635 TATATGACATCCAAAGATCAGTTCAGTCTGTGTGAAACCC-----AAGTCATGTTT 1688
OY 298 GlnPro-----CysPro 301
Db 1689 CACCCCTCAAGTTTGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1748
OY 302 TyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIleLeuLys 321
Db 1749 GATCTGTGATCACTGGAATAATGACACGCCACA---CTTCATGACAAAGTATTT----- 1799
OY 322 AspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlyHisLeuProLeu 341
Db 1800 GATGGTTCATCTGCTGTTCTTTATAGGTACAGCTGTGAGGTTCTCCCTACCTCAGT 1859
OY 342 LysSerPheThrAlaValCysGlnLysAspGlySerTyrAspArgProMetProAlaCys 361
Db 1860 GGAACCTGAACCTGAACCTGATACAAATGACATGAGGCTGCTCTTAACCTCAGTGT 1919
OY 362 ---SerIleValAspCys---GlyProProAspAspLeuProSer---GlyArgValGlu 378
Db 1920 ATCAAGTCACTACCTGCGCAAAACCCCTP-----GTACATCATATATGTTGTGTGAA 1973
OY 379 TyrIleThrGlyProGlyValThr----- 387
Db 1974 ---ATCAAAACCCCAAGTGTGACAAACCTGATAGTCGTGTGGTACCTTTCTTGAGG 2030
OY 387 ----- 387
Db 2031 TTGCACAGGTTACCCCTCCCATTAAGCTAGAGCAGCAAAACCTCTCCAAACCTAGATCC 2090
OY 387 ----- 387
Db 2091 TCACAACCTCTACTGTGACGTTGGCTTTAAAGTTAAACTACCTGAAGGTCAATTACCGG 2150

Oy 388 -----TyrLysAlaValIleGlnTyrSerCysGluGlnThrPheThrThrtMetLysVal 405
 :::|||||: |||||:||||| ::::||| :
 Db 2151 GTAGGGCTCGAGCCATC-----TACAGTCGGAATCGAGACTACTGCACACTTGTGCA 2204
 Oy 406 AsnAspGlyLysTyrValCysGluAlaAspGlyPheTrpThrSerSerLysGlyLys 425
 ::: ||||: |||||:||||| ||||: :
 Db 2205 TCTCAAGCACAGAACA---TGTAAGCTCATTAAGAACTGGAGTGGTGGCCAGCAGCAGC--- 2258
 Oy 426 SerLeuProValCysGluProValCysGlyLeuSerAlaArgThrThrGlyLysArglle 445
 ||| ||||| ||||| |||| | ||
 Db 2259 -----TGTATTCCAGTTTGGGCGGTGCAGACTCTCCCTGCTTCCCTTTTATC 2306
 Oy 446 TyrGlyGlyGlnLysAlaLysProGlyAspPheProTrpGlnValLeuLeu----- 462
 ||| ::: |||||:||||| |||
 Db 2307 TGGAAATGGGAATTTCTACAGAAATAAGCTCACTGGCCCTGGGAGGCCAGAAATCTCTAATGG 2366
 Oy 463 -----LeuGlyGlyThrThrAlaAlaGlyAlaLeuLeuLysAspAsnTrp 477
 ::: ::: |||||:||||| |||
 Db 2367 CTTCAGAACCCAAATATGTGGTTTCTCCCACTGGGAGAGATTCCTATTAAGTAGAANAATGG 2426
 Oy 478 ValLeuThrAlaAlaHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAsp 497
 :::::||||||| |||| | ::: ||||: ||||
 Db 427 ATCGCACTGCTGCCACTGTGTC-----ACCTACTCTGCTACTGCTGAGATTATTT 2477
 Oy 498 ArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrpSer----- 514
 ||| |||| | ::: |||
 Db 2478 GACCCCAATCAGTTTAATAATGTATCTGGGCCAACTACTACCCTGATGATGACGTAGAGACAT 2537
 Oy 515 -----GlnAlaValPheLeuHISglGlyTyrThrHisAspAlaGly 528
 ||||| ||||| ::: |||| | |||
 Db 2538 GACTATGTACAAATGTAAGAGAGGCTCTGAGATCCAGCTGATGATATCTTACACAGCCCCGC 2597
 Oy 529 PheAspAsn---AspIleAlaLeuLysLeuAsnAsnLysValVallLeuSerAsn 547
 ||| ||||| ||||| ||||| ||| ::: :
 Db 2598 AATCTCAACTTTGCAATAGAGCCCTAATTCACACTGAATAACTCTGTTACTTTGCAACACAGA 2657
 Oy 548 IleThrProIleCysLeuProArgLys---GlnAlaGluSerPheMetArgThrAsp 566
 ::: ||||| ||||| ||||| ::: :
 Db 2658 GTCCAAACAATCTGTGCTCTACTGACATTCACAACAAGAAACACTGTGAAGAGGGAACA 2717
 Oy 567 IleGlyThrAlaSerGlyTyrGlyLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMet 586
 ::: ||||| ||||| ||||| ::: :
 Db 2718 TTACAGAGTGTGACAGGTGTGGGGTTTGAATGAAGAACAACACCTATTGCAAGACGATTCAA 2777
 Oy 587 TyrValAspIleProIleValAspHisSglnLysCysThrAlaAlaTyrGlyLysProPro 606
 ::: ||||: ||||| ||| |||: ||||: :
 Db 2778 CAAGCTGTGCTACTCTGTTGTTCAGCCAGCAGCACTGTGAAGAGGGGTACAAAGAACACAGAC 2837
 Oy 607 TyrProArgGlySerValIThrAlaAsnMetLeuCysAlaGlyLeuGluSerLysLys 626
 ||| ||||| ||||| ||||| ||||| ::: ||
 Db 838 TTACCACTG---ACAGTAACAGAAACATCTGTTGTGGAGTTTACAGAAAGGAGCTTAT 2894
 Oy 627 AspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeu---AspSerLthrGlu 645
 ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2895 GATGCTCTCAATGGGAGCACTGGAGGACCTTTAGTGTTCGTGATGATTCGCCGTGACCAA 2954
 Oy 646 ---ArgTrpPheValGlyGlyIleValSerTrpGlySer---MetAsnCysGlyGlnAla 663
 ||||| ::: ||||| ||||| ||||| ||||| ||||| ::: ||||
 Db 2955 AGGGCGTGGCTCTTGAAGAGGATTTGTACGTGGGGCAGTCCAGATGATGTGCGCAAGCGC 3014
 Oy 664 GlyGlnTyrGlyValIYrThrLysValIleAsnTyrIleProTrpIleGluAsnIlelle 683
 ||||| ||||| ||||| ||||| ::: ||||| |||
 Db 3015 AACCAATACGGGGCTTCACTAAAGTAAACGTTTCTGTCTATGTGATGAGCGATTTCAATT 3074

RESULT 8
 US-08-296-014A-1
 ; Sequence 1, Application US/08296014A
 ; Patent No. 5716834
 ; GENERAL INFORMATION:
 ; APPLICANT: Ding, Jeak Ling
 ; APPLICANT: HO, Bow
 ; TITLE OF INVENTION: The Cloned Factor c DNA of the
 ; TITLE OF INVENTION: Singapore Horseshoe Crab, Carlinoscorprius

```

1  ? NUMBER OF INVENTION: rotundicauda and Purification of Factor C Proenzym
2  ? NUMBER OF SEQUENCES: 4
3  ? CORRESPONDENCE ADDRESS:
4  ? ADDRESSEE: Birch, Stewart, Kolach & Birch
5  ? STREET: 8110 Gatehouse Road, Suite 500 East
6  ? CITY: Falls Church
7  ? STATE: Virginia
8  ? COUNTRY: USA
9  ? ZIP: 22042
10 ? COMPUTER READABLE FORM:
11 ? MEDIUM TYPE: Floppy disk
12 ? COMPUTER: IBM PC compatible
13 ? OPERATING SYSTEM: PC-DOS/MS-DOS
14 ? SOFTWARE: PatentIn Release #1.0, Version #1.25
15 ? CURRENT APPLICATION DATA:
16 ? APPLICATION NUMBER: US/08/296,014A
17 ? FILING DATE:
18 ? CLASSIFICATION: 435
19 ? ATTORNEY/AGENT INFORMATION:
20 ? NAME: Murphy, Jr., Gerald M.
21 ? REGISTRATION NUMBER: 28,977
22 ? REFERENCE/DOCKET NUMBER: 1781-105P
23 ? TELECOMMUNICATION INFORMATION:
24 ? TELEPHONE: (703) 205-8000
25 ? TELEFAX: (703) 205-8050
26 ?
27 ? INFORMATION FOR SEQ ID NO: 1:
28 ? SEQUENCE CHARACTERISTICS:
29 ? LENGTH: 4182 base pairs
30 ? TYPE: nucleic acid
31 ? STRANDEDNESS: single
32 ? TOPOLOGY: both
33 ? MOLECULE TYPE: cDNA
34 ? HYPOTHEetical: NO
35 ? ANTI-SENSE: NO
36 ? ORIGINAL SOURCE:
37 ? ORGANISM: Carcinoscorpilus rotundicauda
38 ?
39 ? FEATURE:
40 ? NAME/KEY: CDS
41 ? LOCATION: 569..3817
42 ?
43 ? US-08-296-014A-1
44
45 Alignment Scores:
46 Pred. No.: 4.73e-38 Length: 4182
47 Score: 470.50 Matches: 151
48 Percent Similarity: 39.63% Conservative: 63
49 Best Local Similarity: 27.96% Mismatches: 185
50 Query Match: 12.52% Indels: 141
51 DB: 1 Gaps: 26
52
53 US-09-874-198-2 (1-666) x US-08-296-014A-1 (1-4182)
54
55 Oy 265 HisArgIleGluThr-----LysSerAsnThrVal 274
56 |||||:::|||||:::||||| |||
57 Db 2258 CACAGACACAGACGTCGACAGACCCCTTATTTTGGCAGCTTATGAGCTAGTACTATGTGGTT 2317
58
59 Oy 275 -----ThrIleThrPhe--ValThrAspGluSerGlyAspHisThrGly----- 288
60 ::||| |||:: ||| |||:: |||
61 Db 2318 CTGAATGATTAACCTAACATATTCTGGGCTCTGGCGAACCTGGAATGAAGAACTAGCTGTGA 2377
62
63 Oy 289 -----TrpLysIleHisIstYrThrSerThrAla 297
64 ||||| ||||| |||
65 Db 2378 TATATGGACATCAAGATCAAGTTCAGTTCAGTCTGTGTGGAAAC-----AAGTCATGTTT 2431
66
67 Oy 298 GlnPro-----CysPro 301
68 ||||| |||
69 Db 2432 CAGCCCTCAAGCTTTTGCTTCGATGATGATGATCTGTCAAGACGAAATTAAGCCCAATTCGAT 2491
70
71 Oy 302 TyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIleLeuLys 321
72 |||::: ||||| |||::: |||
73 Db 2492 GATCTCGATATCAGCGAAATGACACGCCACA---CTTATGACACAAGATATT----- 2542
74
75 Oy 322 AsperPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlyHisLeuProLeu 341

```

```

Db 2543 GAGGGCTTATGCTGTTCTTCTATAGGTACAGCTGTGAGGTTCTCCACTACCTCAGT 2602
      |||
      |||
      |||
Qy 342 LysSerPheThrAlaValCysGlnLysAspGlySerTrpAspArgProMetProAlaCys 361
      |||
      |||
      |||
Db 2603 GGAACGAAACCGTAACTGTGTACACAAATGGACATGGAGTCTCTAAACCTGCATGT 2662
      |||
      |||
      |||
Qy 362 ---SerIleValAspCys---GlyProProAspPheProSer---GlyArgValGlu 378
      |||
      |||
      |||
Db 2663 ATCAAGTGTATCATCCCTGCCAATAACCCCTCTCTACATCATATGTCTGTGTGAA 2716
      |||
      |||
      |||
Qy 379 TyrIleThrGlyProGlyValThr----- 387
      |||
      |||
      |||
Db 2717 ---ATCAAAACCCCAAGTGGACAAACATCGATTAAGTGTGTGTGACCTTTCTTGAGG 2773
      |||
      |||
      |||
Qy 387 ----- 387
      |||
      |||
      |||
Db 2774 TTGGCAGCGTTACCCCTCCATTAGCCAGACGCAACCTCTCCAAACCTAGATCC 2833
      |||
      |||
      |||
Db 2834 TCACAACCTCTACTGTGACTTGCTCTAAAGTTAACTACTGAGTCTATTACGG 2893
      |||
      |||
      |||
Qy 388 -----TyrLysAlaValIleGlnTyrSerCysGlnGluThrPheTyrThrMetLysVal 405
      |||
      |||
      |||
Db 2894 GTAGGGTCTCGAGCCATT-----TACAGTGGAGTGCAGATACTACGAACTACTTGA 2947
      |||
      |||
      |||
Qy 406 AsnAspGlyLysTyrValLysGluAlaAspGlyPheTrpThrSerSerLysGlyLys 425
      |||
      |||
      |||
Db 2948 TCTCAGGACAGAGA--TGTGACTTAATGAACTGAGTGTGGCCAGCGAGAC-- 3001
      |||
      |||
      |||
Qy 426 SerLeuProValCysGluProValCysGlyLeuSerAlaArgThrThrGlyArgGle 445
      |||
      |||
      |||
Db 3002 -----TGTATTCAGTTGTGTGACGGTGCAGACTCTCTCTCTCTCTTTTATC 3049
      |||
      |||
      |||
Qy 446 TyrGlyGlyGlnLysAlaLysProGlyAspPheProTrpGlnValLeuIle----- 462
      |||
      |||
      |||
Db 3050 TGGAAATGGAAATCTACAAATAATAGCTAGTGCCTGGCAGGACAGAACTCTAGATGC 3109
      |||
      |||
      |||
Qy 463 -----LeuGlyGlyThrThrAlaAlaGlyAlaLeuLysThrAspAsnTrp 477
      |||
      |||
      |||
Db 3110 CTTGCGACACACATATGTGGCTTCTCCAGTGTGGAGAGTCTTATTTAGTGAATGG 3169
      |||
      |||
      |||
Qy 478 ValLeuThrAlaAlaAlaAlaValTyrGlnGlnLysHisAspAlaSerAlaLeuAspIle 497
      |||
      |||
      |||
Db 3170 ATCGTACGTGTCGCGCAGTGTGC-----ACCTACTCTCTCTCTCTGTGATTTAT 3220
      |||
      |||
      |||
Qy 498 ArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrpSer----- 514
      |||
      |||
      |||
Db 3221 GACCCCAATCAGTTTAAATGTATCTGGGCAAGTACTACCTGATGACAGTAAGACAT 3280
      |||
      |||
      |||
Qy 515 -----GluAlaValPheIleHisGluGlyTyrThrHisAspAlaGly 528
      |||
      |||
      |||
Db 3281 GACTATGTACAGTAAGAGAGGCTCTTGAGATCCAGCTAATCTTAACGACCCGGC 3340
      |||
      |||
      |||
Qy 529 PheAspAsn---AspIleAlaLeuIleLysLeuAsnAsnLysValAlaIleAsnSerAsn 547
      |||
      |||
      |||
Db 3341 AATCTCACTTTGACATAGCCCTTAATCACTGAATAACCTCTGTCTTGACAAACACA 3400
      |||
      |||
      |||
Qy 548 IleThrProIleCysLeuProArgLys---GluAlaGlySerPheMetArgThrAspAsp 566
      |||
      |||
      |||
Db 3401 GTCCAAACCAATCTGTCTCTACTGACATCACAAACAGAACACTTGAAGAGGAGACA 3460
      |||
      |||
      |||
Qy 567 IleGlyThrAlaSerGlyTrpGlyLeuThrGlnArgGlyPheLeuAlaArgAsnLeuMet 586
      |||
      |||
      |||
Db 3461 TTAGCAGTGTACACAGTTGGGGTTTGAATGAATAACACACCTATTTCAGAGCATTA 3520
      |||
      |||
      |||
Qy 587 TyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGlnLysProPro 606
      |||
      |||
      |||
Db 3521 CAAGTGTCTGTCTGTGTGACAGCCACACCTGTGAAGAGGGGTACAAAGAGAGAGAC 3580
      |||
      |||
      |||
Qy 607 TyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGlnSerGlyLys 626
      |||
      |||
      |||

```

```

Db 3581 TTACACATG---ACATACAGAGAACATGTCTGTGACAGTTTACAAAGAGGACGTTAT 3637
      |||
      |||
      |||
Qy 627 AspSerCysArgGlyAspSerGlyValAlaLeuValPheLeu---AspSerGluThrGlu 645
      |||
      |||
      |||
Db 3638 GATGCTGTGAGTGGAGACGTGGAGACCTTTAGTGTGTGTGATGATTCCTCCGCGAA 3697
      |||
      |||
      |||
Qy 646 ---ArgTrpPheValAlaGlyIleValaSerTrpGlySer---MetAsnCysGlyGluAla 663
      |||
      |||
      |||
Db 3698 AGCGGGTGGGTCTTGAAGAGGATGTACGTGGCGCAGTCCAGTGGATGTGCGAAGCGC 3757
      |||
      |||
      |||
Qy 664 GlyGlnTyrGlyValIleThrLysValIleAsnTyrIleProTrpIleGluAsnIleIle 683
      |||
      |||
      |||
Db 3758 AACCACTAGCGGGGCTTCACTAAAGTTAACTGTTTCTGTCTCATGATTAAGCAGTTCA 3817
      |||
      |||
      |||

```

RESULT 9

```

US-08-596-405-1
; Sequence 1, Application US/08596405
; Patent No. 5858706
; GENERAL INFORMATION:
; APPLICANT: Ding, Jeak Ling
; APPLICANT: HO, Bow
; TITLE OF INVENTION: The Cloned Factor C cDNA of the
; TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinus scorpilis
; TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,405
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1781-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ. ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Carcinus scorpilis rotundicauda
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 569..3817
; US-08-596-405-1

```

Alignment Scores:

```

Pred. No.: 4,73e-38
Score: 470.50
Percent Similarity: 39.63%
Best Local Similarity: 27.96%
Query Match: 12,528
DB: 2
Gaps: 26

```



```

; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Carcinoscorpius rotundicauda
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 569..3817
; US-08-877-620-1
;
; Alignment Scores:
; No.: 4,73e-38 Length: 4182
; Percent Similarity: 470.50 Matches: 151
; Best Local Similarity: 39.63% Conservative: 63
; Query Match: 27.96% Mismatches: 185
; DB: 12.52% Indels: 141
; Gaps: 26
;
; US-09-874-198-2 (1-686) x US-08-877-620-1 (1-4182)
;
; 265 HISATGILGLUthr-----LysSerAsnThrVal 274
; 2258 CACAGCTAGATCGTGAAGCCCTTATTGGAGCTAATGATGCTGATGCTGTT 2317
; 275 -----ThrIlePhe---ValThrAspGluSerGlyAspHisIleArgly----- 288
; 2318 CTGAATGATAACCTACATTTCTGGGCTCTGGGAACTGGAAATAAACTGATGTA 2377
; 289 -----TrrLysIleHisIleThrSerThrAla 297
; 2378 TATATGACATCCAGATCAGTTGCGTGTGTGTAAGAAC-----AAGTATGTTT 2431
; 298 GlnPro-----CysPro 301
; 2432 CAGCCCTCAAGTTTGGCTTGCATGATGATCTGTCAAGACAAATTAAGCCAAATGCGAT 2491
; 302 TyrProMetAlaProProAsnGlyHisValSerProValGlnAlaIleLysTyrIleLeuLys 321
; 2492 GATCTGATGATCGTGAATGACAGCCACAC-----CTTCATGAGACAAAGTATT----- 2542
; 322 AspSerPheSerIlePheCysGluThrGlyTyrGluLeuGlnGlnHisLeuProLeu 341
; 2543 GATGGTTCATGCTGCTTCTCTATAGGTACAGCTGAGAGTTTCCACTACCTCAGT 2602
; 342 LysSerPheThrAlaValLysGlnLysAspGlySerTrrAspArgProMetProAlaCys 361
; 2603 GGAACCTGAACGTAACCTTGTACAAACAAATGGACATGAGTGCCTCTAAACCTCATGT 2662
; 362 ---SerIleValAspCys---GlyProProAspPheProSer---GlyArgValGlu 378
; 2663 ATCAAAAGTATACCTGACCCCAAAACCCCT-----GTACATCATATAGTGTGTGGAA 2716
; 379 TyrIleThrGlyProGlyValThrThr----- 387
; 2717 ---ATCAAAACCCCAAGTGGAGCAAACTCGATTAAGTGTGTGGTGCACCTTCTTGAGG 2773
; 387 ----- 387
; 2774 TTGCCACGGTTAACCCCTCCCAATTAGCCAGAGACGCCAAACCTCCTCAAAACCTAGATCC 2833
; 387 ----- 387
; 2834 TCACAACCTCTACTGTGACCTTGGCTTCTTAAAGTTAAACTACTCTAAGTCAATTACCGG 2893
; 388 -----TyrLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMetLysVal 405

```

```

; DB 2894 GTAGGCTCTGAGCCATT-----TACAGTGGGATCGAGATACACTACACTTGGGA 2947
; 406 AsnAspGlyLysTyrValLysGlnAlaAspGlyPheTrrPheSerSerLysGlyLys 425
; 2948 TCTCAAGGCAAGAAG---TGTGACTTAATGGAACCTGAGTGGTGGCCAGCCAGC--- 3001
; 426 SerLeuProValCysGluProValLysGlyLeuSerAlaArgThrIleGlyAlaGlie 445
; 3002 -----TGTATTCAGTTTGTGGACGGTGCAGCTCTCTGCTCTCTTTATC 3049
; 446 TyrGlyGlyGlnLysAlaLysProGlyAspPheProTrrPheIleValIle----- 462
; 3050 TGGAAATGGGAATTCATCAGAAATAGCTAGTGGCCGTGGCAGCCAGAGATCTAGATGG 3109
; 463 -----LeuGlyGlyThrIleAlaIleGlyAlaLeuLeuTyrAspAsnTrr 477
; 3110 CTTCAGACCCACATATGTGTTCTTCCAGTGTGGAGATCTCTATTAATGAAGAAATGG 3169
; 478 ValLeuThrAlaIleHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAspIle 497
; 3170 ATGCTACTGTGCTGCCACTGTGTC-----ACCTACTCTGTACTGTGCTGAGATTATT 3220
; 498 ArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaIleTrrSer----- 514
; 3221 GACCCCAATCAGTTAAATGTATCTGGGCAAGTACTACCGTGTATGACAGTACAGACAT 3280
; 515 -----GluAlaValPheIleHisGluGlyTyrThrHisAspAlaGly 528
; 3281 GACTATGTACAAAGTAAAGAGGCTTGTGAGATCAGCTGATCTTAATCAGACCCCGG 3340
; 529 PheAspAsn---AspIleAlaLeuIleLysLeuAsnAlaLysValIleAsnSerAsn 547
; 3341 AATCTCACTTGTACATAGCCCTTAATTCACATGAACCTCTCTTACTTGTGACACACGA 3400
; 548 IleThrProIleCysLeuProArgLys---GluAlaGluSerPheMetArgThrAsp 566
; 3401 GTCCAAACCAATCTGTGCTGCTACTGACATCACAACAAGAACACTTGAAGAGGGAACA 3460
; 567 IleGlyThrAlaSerGlyTrrPheLysThrGlnArgGlyPheLeuAlaArgAsnLeuMet 586
; 3461 TTAGCAGTGTGACAGTGTGGGTGTGAATGAACAAACACCTATTTCAGAGACATTCAA 3520
; 587 TyrValAspIleProIleValAspHisGlnLysCysThrAlaIleArgLysProPro 606
; 3521 CAAGCTGTGCTACTCTGTGTGTGCAGCCAGCCTGTGAAGAGGGGTACAAAGAGCAGAC 3580
; 607 TyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyLys 626
; 3581 TTACACATG---ACAGTAAACAGAAACATGTCTGTGACAGTTTACAAGAAAGGACGTTAT 3637
; 627 AspSerCysArgGlyAspSerGlyAlaLeuValPheLeu---AspSerGluThrGlu 645
; 3638 GATGCCCTGAGTGTGGAGACAGTGGAGACCTTTAGTGTTCGTGATTTCCCTACCGAA 3697
; 646 ---ArgTrrPheValGlyGlyIleValSerTrrPheGlySer---MetAsnCysGlyAla 663
; 3698 AGCGGTGGGTCTTGAAGAGGATTTGACGTGGGCAAGTCCAGTGATGGCAAGGCG 3757
; 664 GlyIleTyrGlyValTyrThrLysValIleAsnTyrIleProTrrPheIleGluAsnIle 683
; 3758 AACCACTAGCGGGGCTTCACTAAAGTTTACGTTTCTCTGATCAGATTAAGGCAAGTTCAAT 3817
;
; RESULT 11
; US-09-715-994-1
; Sequence 1, Application US/09715994
; Patent No. 6423526
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Human Serine Protease
; FILE REFERENCE: 99-88
; CURRENT APPLICATION NUMBER: US/09/715,994
; CURRENT FILING DATE: 2000-11-17

```

```
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(807)
US-09-715-994-1

Alignment Scores:
Pred. No.: 9,75e-37 Length: 807
Score: 446.50 Matches: 100
Percent Similarity: 53.01% Conservative: 41
Best Local Similarity: 37.59% Mismatches: 92
Query Match: 11.88% Indels: 33
DB: 4 Gaps: 8

-874-198-2 (1-686) x US-09-715-994-1 (1-807)
444 ArgIleTyrGlyGlnLysAlaLysProGlyAspPheProTrpGlnValLeuIleLeu 463
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 AGGATCATTTGGGGCGAAATGCTGAGCTGCTCTCCCGGCGAGCCCTGATAGTG 60

464 GlyGlyThrAla-----AlaGlyAlaLeuTyr 474
||||| |||||||
61 GTGGAGAGACCTTCGAGAGTCCCAATGACAACTGGTTGGAGTGGGGCCCTGCTCT 120

475 AspAsnTrpValLeuThrAlaAlaHisAlaValTyrGlnLysHisAlaSerAla 494
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
121 GCCTCCGTGATCTCCACAGACGCTCATGTGCTCCGCTCCCGGTAGAGACACCGG 180

495 LeuAspIle---ArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrp 513
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
181 ATACAGATGCTCCAGAGGATGTACCCGCTGCTGCTGCTGATGATGTGCGACAA 240

514 SerGlnAla-----ValPheIleHisGlnGlyTyrThrHisAsp 526
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
241 TCGGGGCGAGTCAACAGCTCAGCTGCCCGGAGTGGTGTCCACCCAGACTTC---A 297

527 AlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnLysValValIleAsnSer 546
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
298 CAAACTACACACGATATAGCTGTGTCAGCTGACGAGGAGCGCTGGCCCGGAGCC 357

547 AsnIleThrProIleCysLeuProArgLysGlnAlaGlnSerPheMetArgThrAsp 566
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
358 CAGTTATGCTGTCTGCTGCTGCCCAAGCTTGAGCCCTGAAAGC---CCGGCCCC 414

567 IleGlyThrAlaSerGlyTyrGlyLeuThrGlnArgGlyPhe-----580
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
415 CTGGCCCTGGTGGCGGCTGGGCGCATCTCCATCCCAATGTGACAGTGATGATCAT 474

581 -----LeuAlaArgAsnLeuMetLysTyrValAspIleProIleValAsp 594
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
475 AGCAGTGGCACACGAGCTGTGATGATGTCCTGACATGTCAGATGTCAGTCCGG 534

595 HisGlnLysCysThrAlaAlaTyrGlnLysProProTyrProArgLysSerValThrAla 614
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
535 CAGCGTGAAGGCAAACTAGACTATGATCCCGCTG---GGCAATTCAGACGCTCAC 591

615 AsnMetLeuCysAlaLysLeuGlnSerGlyLysAspSerCysArgGlyAspSerGly 634
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
592 AACATGTTCTGCTGCTGCTACTACGAGGGGGGCAAGACAGTGGCTTGAGATAGG 651

635 GlyAlaLeuValPheLeuAspSerGlnThrGlnArgTrpPheValGlyGlyIleValSer 654
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
652 GGGGCTTTGCTCATCTTGTGACTTGAGCCACGCGTGGTGTGCAAGCGCTGTCTCC 711

655 TrpGlnSer---MetAsnCysGlnGlnAlaGlyGlnTyrGlyValTyrThrLysValIle 673
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
712 TGGGGGGGACCTGAAGATGGCGGACAGACAGCTTATGAGGCTTACACAAAGGCTCC 771
```

```
QY 674 AsnTyrIleProTrpIle 679
||||| |||||||
DB 772 AATTACGTGAGACTGGGTG 789

RESULT 12
US-09-027-337-9
; Sequence 9, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 9
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SMC19 mRNA sequence (U20428)
US-09-027-337-9

Alignment Scores:
Pred. No.: 2,18e-34 Length: 2900
Score: 432.50 Matches: 204
Percent Similarity: 39.58% Conservative: 94
Best Local Similarity: 27.09% Mismatches: 285
Query Match: 11.51% Indels: 174
DB: 2 Gaps: 37

US-09-874-198-2 (1-686) x US-09-027-337-9 (1-2900)
QY 13 SerValAlaThrProLeuGlyProLysTrpProGlnProValPheGlyArgLeuAlaSer 32
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
374 AGTCATGCTGCTGCCCGCGGCGC---GGCTCCCTGAAAGTCTTGTGTGTCACCTC---AG 427

33 ProGlyPheProGlyGlnTyrAlaAsnAspGlnArgArgTrpThrLeuThrAlaPro 52
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
428 GGTGGCTTCTCCACGAGCTCCCAAAACAGT-----ACAGAGGAC 466

53 ProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuGlnLeuSerHisLeuCys 72
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
467 CCAGGACAAACAGCTGAGCTTGCTGCACGCGCGGTGTGAGGCTG-----514

73 GlnTyrAspPheValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuCysGlyGln 92
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
515 -----ATGGGCTTACACGACGCGCGCTTCCCTGCACAGCCCTACCCCGCTCATG 562

93 GlnSerThrAspThrGlnArgAlaProGlyLysAspThrPheTyrSerLeuGlySerSer 112
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
563 CCGGCTGCCAGTGGCTGCGG-----GGACGGACGACGAGTGTGAGCTACTGAGACTGA 616

617 CTGCGAGCTTGACTGCGGCTCGCAC-----640
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
113 LeuAspIleThrPheArgSerAspTyrSerAsnGlnLysProPheThrGlyPheGlnAla 132
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
617 CTGCGAGCTTGACTGCGGCTCGCAC-----640

133 PheTyrAlaAlaLysPheLysPheGlnVal-----Ala 145
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
641 ---GACCGCGGACGACCTGCTGAGCTGTACAAACCTGAGCCCATGAGCCCGCCACG 697

146 ProGlyGlu-----AlaProThrCysAsp-----153
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
698 CTTGTGAGTGTGTGGCAGCTTACCTCCCTCTACAACTGACCTTCCACTCCCTCCAC 757

154 -----HisIscysHisAsn---HisLeuGlyGlyPheTyrCysSerCysArg 168
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
758 GACGCTGCTGCTGCTGCTGCTGATTAACAACTGACGCGGCGGCTTGGAG-GCC 816

169 AlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSerGlyGlnVal 188
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```

Db      817 ACCTTCTTCAGTCGCTAGG-----ATGAGCAGCTGTGGAGGCCGCTTA 861
Oy      189 PheThrGlnArgSerGlyGluLeuSerSerProGluTyrProArgProTyrProLysLeu 208
Db      862 ----CGTAAAGCCAGGGGACATTCACACAGCCCTACTACCCAGGCCACATCCACCCAAC 918
Oy      209 SerSerCysThrTyrSerIleSerLeuGluGlyPheSerValIleLeuAspPheVal 228
Db      919 ATTAGCTGCACATGGAATATGAGTGCCCAACACACACATGTGAAGGCGCCCTTCANA 978
Oy      229 GluSerPheAspValGluThrHis---ProGluThrLeuCysProTyrTyrAspPheLys 247
Db      979 TTTCTTCTTACTCTGCGAGCCCGCGCTGCGCGCACCTCCGCCCAAGGACTACGTGGAG 1038
Oy      248 IleGlnThrAspArgGluGluHisGlyProPheCysGlyLysThrLeuProHisAsnArgIle 267
Db      1039 ATCAATGGGGAGAAA-----TACTGGGAGAGAGGTCCCGCATTTCTGTCGTC 1083
Oy      268 GluThrLysSerAsnThrValThrIleThrPheValThrAspLeuSerGlyAsnHisThr 287
Db      084 ACCAGCAGACAGCAACAAGATACAGATTCGCTTCACATCAGATCAGTCTACACCCGACAC 1143
Oy      288 GlyThrLysIleHisTyrThrSer-----ThrAlaGlnProCysProTyrTyrProMetAla 305
Db      1144 GGTCTTCTTAGCTGAATATCTCTCTACGACTCCAGTGCAGTCCATGCCCGG-GCAGTTTCAC 1202
Oy      306 ProProAsnGlyHisValSerProValGlnAlaLysTyr----- 318
Db      1203 GTGCCCCAGGGGGCGGTGATCCGGAAGAGCTGCCCTGTGATGGCTGGCGCACCTCACCC 1262
Oy      319 IleLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnGlyHis 338
Db      1263 GACCACAGCAGATAGACTCAACTGCACACTGCAGCGCGGCACAGTTCAAGTGCACAGAG 1322
Oy      339 LeuProLeuLysSerPheThrAlaValLysGlnLys-----AspGlySer 353
Db      1323 AAGTTCTGCAGCTCTTTCGG---GTCTCGCAGACTGTGAACGAGTGCAGCAACACAG 1379
Oy      354 -----TTPAspArgPro 357
Db      1380 GACGAGCAGGTTGCATTTGTCCGGACCCAGACCTTCAGGTGTTCCAAATGGAGTGCCT 1439
Oy      358 Met-----ProAla-CysSer---IleValAspCysGlyPro---ProAspAspLeuP 373
Db      1440 CTCGAAACACCCAGCAGTGCATTCGGAAGGACGACCTGGGAGCGGTGCGACAGGCCCTC 1499
Oy      373 OserGlyArgValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleG 393
Db      500 CTGCCCCCAAGGTGAACGTGCTGCTGCTGTGACCAACACACCC----- 1539
Oy      393 ntyrSerCysGluGluThrPheTyrThrMetLysValAsn-----AspGlyLysTyr 410
Db      1540 -TACCCTCCCTCAATAGGGCTCTGCTTGAGCAAGGCCAACCTGAGTGTGACGGAGGA 1598
Oy      410 rValCysGluAlaAspGlyPheThrPheSerSerLysGlyGlnLysSerLeuProValC 430
Db      1599 GGACTGT---ACGACGCGC-----TCAGATGAGAAGAGAC-----Tg 1631
Oy      430 sGluProValLysGlyLeuSerAlaArgThrThrGlyGluArgIleTyrGlyGlnLys 450
Db      1632 CGAC-----TGTGGCTCTGGGCTCATTCACAGACAGGCTCGTTGTTGGGGCGCGGA 1685
Oy      450 sAlaLysProGlyAspPheProTyrGlnValLeuIle-----LeuGly---GlyThrTh 467
Db      1686 TCGGAGATGAGGCGAGTGGCTCTGGAGGTAAAGCTGCACTGCTTGGCGCAGGGCCACAT 1745
Oy      467 rAlaAlaGlyAlaLeuLeuTyrAspAsnThrValLeuThrAlaAlaAlaValTyrG 487
Db      1746 CTGGCGTCTCTCCATCATCTCCCAATGTGGTGTCTGCGCGACACCTGCAATCGA 1805
Oy      487 uGlnLys----- 489
Db      1806 TGACAGAGGATTTCAGTACTCAGACCCGACGAGAGCGGCTTCCTGGGCTTGCAGAC 1865

```

```

Oy      490 -HisAspAlaSerAlaLeuAspIleLeuMetGlyThrIleLysArgLeuSerProHisTyr 509
Db      1866 AAGGCCAGCGCAGGCCCTGGGGTGCAGAGCCGAGG-CTCAACGCGATCTCTCCAC-- 1922
Oy      509 rThrGlnAlaTrpSerGluAlaValAlaPheIleHisGluGlyTyrThrHisAspAlaGly 529
Db      1923 -----CCCTTTCGAATGACTTCACCTT 1945
Oy      529 eAspAsnAspIleAlaLeuIleLysLeuAsnLysValValIleAsnSerAsnIleTh 549
Db      1946 CGACTATGACATCCGCTGCTGTGAGCTGCAGAAACCGCGAGACTACAGCTCCATGCTGG 2005
Oy      549 rProIleCysLeuProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyTh 569
Db      2006 GCCCATCTGCTCTCCG-----GACGCTTCGCACTGCTTCTCCGCGGCAAGGCCATCTG 2059
Oy      569 rAlaSerGlyTrpGlyLeuThrGlnArgGlyPheLeu---AlaArgAsnLeuMetTyrVa 588
Db      2060 GGTACAGGGGTGGGAGACACCCAGTATGAGGAGCTGGCGCGCTGATCTGCANAAGG 2119
Oy      588 lAspIleProIleValAlaAspHisGlnLysCysThrAlaAlaTyrGlnLysProTyrPr 608
Db      2120 TGAATCCGCGTCAATACACAGACCACTGCGAGAACTCTCCGCCAG----- 2168
Oy      608 oArgGlySerValThrAlaAsnMetLysAlaGlyLeuGluSerGlyLysAspSe 628
Db      2169 -----CAGATACCCCGCGCATGATGTGCGGGCTTCTTCAAGGCGCGCTGACATC 2221
Oy      628 rCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGlyThrGluArgTrpPh 648
Db      2222 CTCGCAAGGGTATTCGCGGGGACCCCTGTCCACAGCGTGGAGCGGAT---GGCGGATCTT 2278
Oy      648 eValGlyGlyIleValSerTrpGlySerMetAsnCysGlyGlnAlaGlyGlnTyrGlyVa 668
Db      2279 CCAGGCCGGGTGTGTGAGTGGGGAGA---CCGTCGCTCAGAGGACAAACAGCGCGGT 2334
Oy      668 lTyrThrLysValIleAsnTyrIleProTrpIleGlu 680
Db      2335 GTACACAGAGCTCCTCTGTGTTGCGGAATGATCAAA 2371

RESULT 13
US-09-644-600-9
; Sequence 9, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: TadG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CTP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 9
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SNC-19; GeneBank Accession No. 6451500 #U20428
US-09-644-600-9

Alignment Scores:
Pred. No.: 2,18e-34
Score: 432.50
Percent Similarity: 39.58
Best Local Similarity: 27.09%
Query Match: 11.51%
DB: 4
Matches: 2900
Conservative: 204
Mismatches: 94
Indels: 174
Gaps: 37

```



```

Oy 374 Ser----- 374
Db 3573 TCATATTGCAGATTTCAGTTTAATTCATTTCCACACTGCATGCATACACATCAACACAGAAGGA 36322
Oy 375 -----GlyArgValGluTyrIleThrGlyProGlyValThrTyrIleValIle 392
Db 3633 TTTCATATTAGATATCAAAACCATAGATATCCAGATACACACATATCAAAAAATTAACAC 36922
Oy 393 GlnTyrSerCysGluGluThr-----PheTyr 401
Db 3693 CAAAACCTCTGTCGAACACAAAGAAATGCATATGAGAGAAACATATTTTTTTA 37522
Oy 402 ThrMetIleValAsnAspGlyLysTyrValCysGluAlaAspGlyPheThrPheSer 421
Db 3753 AACTGAGATA-----TTGCGCAAAATGTTTATACAAAGAGT 37921
Oy 422 LysGlyGluLysSerIleu 427
Db 3792 TTGAACAAATAATCCCTG 3809
Oy 427 ----- 427
Db 3809 ----- 3809
Oy 428 ----- 428
Db 3809 ----- 3809
Oy 429 ----- 429
Db 3809 ----- 3809
Oy 430 ----- 430
Db 3809 ----- 3809
Oy 431 ----- 431
Db 3809 ----- 3809
Oy 432 ----- 432
Db 3809 ----- 3809
Oy 433 ----- 433
Db 3809 ----- 3809
Oy 434 ----- 434
Db 3809 ----- 3809
Oy 435 ----- 435
Db 3809 ----- 3809
Oy 436 ----- 436
Db 3809 ----- 3809
Oy 437 ----- 437
Db 3809 ----- 3809
Oy 438 ----- 438
Db 3809 ----- 3809
Oy 439 ----- 439
Db 3809 ----- 3809
Oy 440 ----- 440
Db 3809 ----- 3809
Oy 441 ----- 441
Db 3809 ----- 3809
Oy 442 ----- 442
Db 3809 ----- 3809
Oy 443 ----- 443
Db 3809 ----- 3809
Oy 444 ----- 444
Db 3809 ----- 3809
Oy 445 ----- 445
Db 3809 ----- 3809
Oy 446 ----- 446
Db 3809 ----- 3809
Oy 447 ----- 447
Db 3809 ----- 3809
Oy 448 ----- 448
Db 3809 ----- 3809
Oy 449 ----- 449
Db 3809 ----- 3809
Oy 450 ----- 450
Db 3809 ----- 3809
Oy 451 ----- 451
Db 3809 ----- 3809
Oy 452 ----- 452
Db 3809 ----- 3809
Oy 453 ----- 453
Db 3809 ----- 3809
Oy 454 ----- 454
Db 3809 ----- 3809
Oy 455 ----- 455
Db 3809 ----- 3809
Oy 456 ----- 456
Db 3809 ----- 3809
Oy 457 ----- 457
Db 3809 ----- 3809
Oy 458 ----- 458
Db 3809 ----- 3809
Oy 459 ----- 459
Db 3809 ----- 3809
Oy 460 ----- 460
Db 3809 ----- 3809
Oy 461 ----- 461
Db 3809 ----- 3809
Oy 462 ----- 462
Db 3809 ----- 3809
Oy 463 ----- 463
Db 3809 ----- 3809
Oy 464 ----- 464
Db 3809 ----- 3809
Oy 465 ----- 465
Db 3809 ----- 3809
Oy 466 ----- 466
Db 3809 ----- 3809
Oy 467 ----- 467
Db 3809 ----- 3809
Oy 468 ----- 468
Db 3809 ----- 3809
Oy 469 ----- 469
Db 3809 ----- 3809
Oy 470 ----- 470
Db 3809 ----- 3809
Oy 471 ----- 471
Db 3809 ----- 3809
Oy 472 ----- 472
Db 3809 ----- 3809
Oy 473 ----- 473
Db 3809 ----- 3809
Oy 474 ----- 474
Db 3809 ----- 3809
Oy 475 ----- 475
Db 3809 ----- 3809
Oy 476 ----- 476
Db 3809 ----- 3809
Oy 477 ----- 477
Db 3809 ----- 3809
Oy 478 ----- 478
Db 3809 ----- 3809
Oy 479 ----- 479
Db 3809 ----- 3809
Oy 480 ----- 480
Db 3809 ----- 3809
Oy 481 ----- 481
Db 3809 ----- 3809
Oy 482 ----- 482
Db 3809 ----- 3809
Oy 483 ----- 483
Db 3809 ----- 3809
Oy 484 ----- 484
Db 3809 ----- 3809
Oy 485 ----- 485
Db 3809 ----- 3809
Oy 486 ----- 486
Db 3809 ----- 3809
Oy 487 ----- 487
Db 3809 ----- 3809
Oy 488 ----- 488
Db 3809 ----- 3809
Oy 489 ----- 489
Db 3809 ----- 3809
Oy 490 ----- 490
Db 3809 ----- 3809
Oy 491 ----- 491
Db 3809 ----- 3809
Oy 492 ----- 492
Db 3809 ----- 3809
Oy 493 ----- 493
Db 3809 ----- 3809
Oy 494 ----- 494
Db 3809 ----- 3809
Oy 495 ----- 495
Db 3809 ----- 3809
Oy 496 ----- 496
Db 3809 ----- 3809
Oy 497 ----- 497
Db 3809 ----- 3809
Oy 498 ----- 498
Db 3809 ----- 3809
Oy 499 ----- 499
Db 3809 ----- 3809
Oy 500 ----- 500
Db 3809 ----- 3809
Oy 501 ----- 501
Db 3809 ----- 3809
Oy 502 ----- 502
Db 3809 ----- 3809
Oy 503 ----- 503
Db 3809 ----- 3809
Oy 504 ----- 504
Db 3809 ----- 3809
Oy 505 ----- 505
Db 3809 ----- 3809
Oy 506 ----- 506
Db 3809 ----- 3809
Oy 507 ----- 507
Db 3809 ----- 3809
Oy 508 ----- 508
Db 3809 ----- 3809
Oy 509 ----- 509
Db 3809 ----- 3809
Oy 510 ----- 510
Db 3809 ----- 3809
Oy 511 ----- 511
Db 3809 ----- 3809
Oy 512 ----- 512
Db 3809 ----- 3809
Oy 513 ----- 513
Db 3809 ----- 3809
Oy 514 ----- 514
Db 3809 ----- 3809
Oy 515 ----- 515
Db 3809 ----- 3809
Oy 516 ----- 516
Db 3809 ----- 3809
Oy 517 ----- 517
Db 3809 ----- 3809
Oy 518 ----- 518
Db 3809 ----- 3809
Oy 519 ----- 519
Db 3809 ----- 3809
Oy 520 ----- 520
Db 3809 ----- 3809
Oy 521 ----- 521
Db 3809 ----- 3809
Oy 522 ----- 522
Db 3809 ----- 3809
Oy 523 ----- 523
Db 3809 ----- 3809
Oy 524 ----- 524
Db 3809 ----- 3809
Oy 525 ----- 525
Db 3809 ----- 3809
Oy 526 ----- 526
Db 3809 ----- 3809
Oy 527 ----- 527
Db 3809 ----- 3809
Oy 528 ----- 528
Db 3809 ----- 3809
Oy 529 ----- 529
Db 3809 ----- 3809
Oy 530 ----- 530
Db 3809 ----- 3809
Oy 531 ----- 531
Db 3809 ----- 3809
Oy 532 ----- 532
Db 3809 ----- 3809
Oy 533 ----- 533
Db 3809 ----- 3809
Oy 534 ----- 534
Db 3809 ----- 3809
Oy 535 ----- 535
Db 3809 ----- 3809
Oy 536 ----- 536
Db 3809 ----- 3809
Oy 537 ----- 537
Db 3809 ----- 3809
Oy 538 ----- 538
Db 3809 ----- 3809
Oy 539 ----- 539
Db 3809 ----- 3809
Oy 540 ----- 540
Db 3809 ----- 3809
Oy 541 ----- 541
Db 3809 ----- 3809
Oy 542 ----- 542
Db 3809 ----- 3809
Oy 543 ----- 543
Db 3809 ----- 3809
Oy 544 ----- 544
Db 3809 ----- 3809
Oy 545 ----- 545
Db 3809 ----- 3809
Oy 546 ----- 546
Db 3809 ----- 3809
Oy 547 ----- 547
Db 3809 ----- 3809
Oy 548 ----- 548
Db 3809 ----- 3809
Oy 549 ----- 549
Db 3809 ----- 3809
Oy 550 ----- 550
Db 3809 ----- 3809
Oy 551 ----- 551
Db 3809 ----- 3809
Oy 552 ----- 552
Db 3809 ----- 3809
Oy 553 ----- 553
Db 3809 ----- 3809
Oy 554 ----- 554
Db 3809 ----- 3809
Oy 555 ----- 555
Db 3809 ----- 3809
Oy 556 ----- 556
Db 3809 ----- 3809
Oy 557 ----- 557
Db 3809 ----- 3809
Oy 558 ----- 558
Db 3809 ----- 3809
Oy 559 ----- 559
Db 3809 ----- 3809
Oy 560 ----- 560
Db 3809 ----- 3809
Oy 561 ----- 561
Db 3809 ----- 3809
Oy 562 ----- 562
Db 3809 ----- 3809
Oy 563 ----- 563
Db 3809 ----- 3809
Oy 564 ----- 564
Db 3809 ----- 3809
Oy 565 ----- 565
Db 3809 ----- 3809
Oy 566 ----- 566
Db 3809 ----- 3809
Oy 567 ----- 567
Db 3809 ----- 3809
Oy 568 ----- 568
Db 3809 ----- 3809
Oy 569 ----- 569
Db 3809 ----- 3809
Oy 570 ----- 570
Db 3809 ----- 3809
Oy 571 ----- 571
Db 3809 ----- 3809
Oy 572 ----- 572
Db 3809 ----- 3809
Oy 573 ----- 573
Db 3809 ----- 3809
Oy 574 ----- 574
Db 3809 ----- 3809
Oy 575 ----- 575
Db 3809 ----- 3809
Oy 576 ----- 576
Db 3809 ----- 3809
Oy 577 ----- 577
Db 3809 -----
```

Pred. No.:	5_9e-32	Length:	3919
Score:	411.00	Matches:	133
Percent Similarity:	41.15%	Conservative:	67
Best Local Similarity:	27.37%	Mismatches:	168
Query Match:	10.94%	Indels:	118
DB:	2	Gaps:	20
US-09-874-198-2 (1-686) x US-09-021-287-4 (1-3919)			
QY	11 CysglSerValAlaThrProLeuGlyProLysTrpProGluProValaPheGlyArgLeu	30	
DB	2499 TGTGGTGGACTCTTACCAACTTAAC-----GGCCACCATTA	2534	
QY	31 AlaSerProGlyPhePheProGlyGluTrpAlaAsnAspGlnGlaArgGlyPheLeu	50	
DB	2535 ACCACCCCTGGCTGGCCCAAGAGATCCCTCTTAATAAAGACTGTGTGGGAAGTGGTT	2594	
QY	51 AlaProGlyGlyTrpAlaGluArgLeuTrpPheThrHisPheAspSerGluLeuSerHis	70	
DB	2595 GCACCAACCCGATACACAATTTCTGTGAAGTTTGATTTTTTGAAATGGAAAGCAATGAA	2654	
QY	71 LeuGlyGluTrpAspPheValLys-----LeuSerSerGlyAlaLysValLeu	86	
DB	2655 GTTTGCAAAATTCGATTTATTTGGGAGATCTGGAGTGGCTTTCTCTTGAGCTTAAC	2714	
QY	87 AlaThrLeuGlyGlnGlnLeuSerThrAspThrGlnArgAlaProGlyLysAspPhe	106	
DB	2715 GGCAAAATCTGTGGCGCTGAAGTG-----CTGAAGCTATC	2750	
QY	107 TyrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTrpSerAsnGluysPro	126	
DB	2751 ACATCCCAAGTTCAACAAATATGAGATTTGAATTCGAATTCGACAAATACGTATTC	2807	
QY	127 PheThrGlyPheGlnAlaPheTrpAlaAlaGluAspIleAspGluCysGlnValaLarPro	146	
DB	2808 --AAGCGTTTCAAGACAACTTTTTTCTCA--GACAAGAGATGAATGCTTAAGAGTAAT	2861	
QY	147 GlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGlyGlyPheTrpCysSer	166	
DB	2862 GGT-----GGATGTCAAGCAGATGTGTCTACACAGATGGGAGAGCTACATGTCTCAA	2912	
QY	167 CysArgAlaGlyTrpValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSergly	186	
DB	2913 TGCCGATATGGATTTGTGTACATGACAAATTAACATGATTTGCAAGAGAGCTAGTGTA	2972	
QY	187 GlnValPheThrGlnArgSerglyGluLeuSerSerProGluTrpProArgProArgPro	206	
DB	2973 CAGAAATCCACAGTCCCAAGTGGCTCATCACAGTCCCACTGCGCAGACAGAAATGACCA	3032	
QY	207 LysLeuSerSerCysThrTrpSerIleSerLeuGlnGluGlyPheSerValIleLeuAsp	226	
DB	3033 AGCAGAAAGATACCTTGTGGAAATACAGGCCACATCCCGCCACCCAGATCAAAATTAACC	3092	
QY	227 PheValGluSerPheAspValGluThrHisProGluThrLeuCysProTrpArgPheLeu	246	
DB	3093 TTTAGTGTA--TTTGAGATTGAGCAGCATCAAGAA-----TGTGCTTAATGACACATTA	3143	
QY	247 LysIleGlnThrAspArgGlnGluHisGlyPro-----PheCysGlyLysThr	262	
DB	3144 GAAGATTTTGTATGAGAAACAACAAGATCCACCATTTTGGACAGCATATGTGGCAACAAG	3203	
QY	263 LeuProHisArgIleGluThrTrpLysSerAsnThrValThrIleThrPheValThrAspGlu	282	
DB	3204 ATACCAAGTCCCTTGTGGCTACTGGAAATTAATTTGTTGCGTTTGTTCATGATCA	3263	
QY	283 SerGlyAspHisThrGlyTrpLysIleHisTrpTrpSer-----Thr	296	
DB	3264 TCTGTCAAAAGAAAGAGCTTTCAAGCCACACATTTACAGAGATGTGGCGAGATTTGAAA	3323	
QY	297 AlaGlnProCysPro-----TyrProMetAlaProProAsnGlyHisVal	311	
DB	3324 GCACAATTCAAAACAGACAGATCTGTACTACATCTCACTGTTGGTATGAACAATCAACCA	3383	

```

Oy 312 SerProValGlnAlaLysrTyrIleLeuLysAsp-----Ser 323
    ||| : : : : : ||| |||
Db 3384 GGACAGGTTGACCTGTGAATGGCTATTAGATACGAACGGGGCTTCGACTTGAAATTAATCC 3443
    ||| : : : : : ||| |||
Oy 324 pheSerIlephe-----CysGluThrGlyTyrGluLeuGlnGlyHis 338
    ||| ||| : : : ||| ||| : : : : :
Db 3444 TTCCAGACATTGGAAGTGGAGAAAGACAGACTGTGGCTATGACTATGTGAGAGCTCTTT 3503
    ||| ||| : : : ||| ||| : : : : :
Oy 339 LeuProLeuLysSerPheThrAlaVal-----CysGlnLysAspGlySer 353
    ||| ||| ||| ||| ||| ||| |||
Db 3504 GATGGCTGTGATTCAC---ACACCTGTGGGGCTGTGTCGATTCGT-----GGATCC 3551
    ||| ||| ||| ||| ||| ||| |||
Oy 354 TrpAspArgPrometProAlaCysSerIleValAspCysGlyProProAspAspLeuPro 373
    3552 -----GGGCGCCACGAAAGAGATTAT 3572
    ||| ||| ||| ||| ||| ||| |||
Oy 374 Ser----- 374
    ||| ||| ||| ||| ||| ||| |||
Db 3573 TCCAATTGGACATTCACTTTAATTCATTTCCACACTGATGACACATCAACAGAGCGCA 36322
    ||| ||| ||| ||| ||| ||| |||
Oy 375 -----GlyArgValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIle 392
    ||| : : : ||| ||| ||| ||| |||
Db 3633 TTTCATATTAGATTGAAACAAAGCATAGATATCCAGATACACACATACCAAAAATATACAC 36922
    ||| ||| ||| ||| ||| ||| |||
Oy 393 GlnTyrSerCysGluGluThr-----PheTyr 401
    ||| ||| ||| ||| ||| ||| |||
Db 3693 CAAACCTCTCTCGAACACAAAGAGATGCTATATGAGAGAGACATATATTTTTTTA 37522
    ||| ||| ||| ||| ||| ||| |||
Oy 402 ThrMetLysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheThrPheSerSer 421
    ||| ||| ||| ||| ||| ||| |||
Db 3753 AAATCGAAGATA-----TTGGACAAATGTTTATACAAAGAGT 3791
    ||| ||| ||| ||| ||| ||| |||
Oy 422 LysGlyLysLeu 427
    ||| ||| ||| ||| ||| ||| |||
Db 3792 TTGAACAAAAAATCCCTG 3809
    ||| ||| ||| ||| ||| ||| |||

RESULT 16
US-09-240-473-4
Sequence 4, Application US/09240473
Patent No. 6297011
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian Tolloid-Like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plackney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/240,473
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296.93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-9160
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3919 base pairs
TYPE: nucleic acid

```

[illegible]


```

Qy 207 LysLeuSerSerCysThrTyrSerIleSerLeuGluGlyPheSerValIleuAsp 226
      |||||::: |||||
Db 2637 AGCAGGAAGAATGCACTGGAAATCAGCCGACCTCTGGCCACCGAATCAATGAGCC 2696
      |||||::: |||||
Qy 227 PheValGluSerPheAspValGluThrHisProGluThrLeuCysProTyrAspPheLeu 246
      |||||::: |||||
Db 2697 TTATAGTAA--TTTGAGATTGAGCAGCATCAAGAA--TGCTCTATGACCACTTA 2747
      |||||::: |||||
Qy 247 LysIleGluThrAspArgGluGluHisGlyPro-----PheCysGlyIleThr 262
      |||||::: |||||
Db 2748 GAAGTATTGATGGAGAAACAGAAAGTCAACCCATTCTTGAGACATAATGGCAACAA 2807
      |||||::: |||||
Qy 263 LeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGlu 282
      |||||::: |||||
Db 2808 ATACCCAGATCCCTCTGCTACTGGAAATTAATGTTTGTGCTTGTTCGTATGCA 2867
      |||||::: |||||
Qy 283 SerGlyAspHisThrGlyTrpLysIleHisTyrThrSer-----Thr 296
      |||||::: |||||
Db 2868 TCTGTTCAAGAAAGAGCTTTCAAGCCACACATTTCAAGAGTGTGGCGAGCATTGAAA 2927
      |||||::: |||||
      297 AlaGluProCysPro-----TyrProMetAlaProProAsnGluHisVal 311
      |||||::: |||||
Db 2928 GCAGATCAAAACCAAGAGATCTGTACTACATGCTCACTTGGTATACAACTACCA 2987
      |||||::: |||||
Qy 312 SerProValGluAlaLysTyrIleLeuLysAsp-----Ser 323
      |||||::: |||||
Db 2988 GCACAGGTGACTGTGAATGCGTATGATGACAGACGGGCTCTGCACTGAATATATCC 3047
      |||||::: |||||
Qy 324 PheSerIlePhe-----CysGluThrGlyTyrGluLeuGluGlnGlyHis 338
      |||||::: |||||
Db 3048 TTCCAGACATTTTGAACTGAGAGAGAGAGCGACTGTGCTATGACTATGTGGAGCTCTTT 3107
      |||||::: |||||
Qy 339 LeuProLeuLysSerPheThrAlaVal-----CysGlnLysAspGlySer 353
      |||||::: |||||
Db 3108 GATGCTCTGATTCA--ACAGCTGTGGGGCTTGGTCGATTCGT-----GGATCC 3155
      |||||::: |||||
Qy 354 TyrAspArgProMetProLacCysSerIleValAspCysGlyProProAspLeuPro 373
      |||||::: |||||
Db 3156 -----GGGCCACACAGAGATTTAT 3176
      |||||::: |||||
Qy 374 Ser----- 374
      |||||::: |||||
Db 3177 TCAATTGGAGATTCACTTTTAATTCATTTCCACACTGATGACAAATCAAGAAAGGA 3236
      |||||::: |||||
Qy 375 -----GlyArgValGluTyrIleThrGlyProGlyValThrTyrLysAlaValIle 392
      |||||::: |||||
Db 3237 TTTCATATAGATATACAAAGCATATATCCAGATACACACATACCAAAAAATTAACAC 3296
      |||||::: |||||
      393 GlnTyrSerCysGluGluThr-----PheTyr 401
      |||||::: |||||
Db 3297 CAAACCTCTGTACAGAACAAAGAAATGTCATATGAGAGAGACATATTTTATTA 3356
      |||||::: |||||
Qy 402 ThrMetLysValIleAsnAspLysTyrValCysGluAlaAspGlyPheTrpHisSer 421
      |||||::: |||||
Db 3357 AAACGTGAAGATA-----TTGGCACAAATGTTTATACAAAGAGT 3395
      |||||::: |||||
Qy 422 LysGlyGluLysSerLeu 427
      |||||::: |||||
Db 3396 TTGACCAAAAATCCCTG 3413
      |||||::: |||||
      RESULT 18
      US-09-432-473-1
      ; Sequence 1, Application US/09432473
      ; Patent No. 6365715
      ; GENERAL INFORMATION:
      ; APPLICANT: ARLETH, ANTHONY J.
      ; APPLICANT: WILLETTTE, ROBERT N.
      ; APPLICANT: ELSHOURBAGY, NABIL A.
      ; TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE PROTEIN
      ; FILE REFERENCE: ATG-50038-D1
      ; CURRENT APPLICATION NUMBER: US/09/432,473
      ; CURRENT FILING DATE: 1999-11-01

```

```

; EARLIER APPLICATION NUMBER: 08/991,408
; EARLIER FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: 60/034,471
; EARLIER FILING DATE: 1997-01-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-432-473-1

Alignment Scores:
  Pred. No.: 9,2e-32      Length: 5145
  Score: 411.00          Matches: 133
  Percent Similarity: 41.15%      Conservative: 67
  Best Local Similarity: 27.37%   Mismatches: 168
  Query Match: 10.94%           Indels: 128
  DB: 4                      Gaps: 20

US-09-874-198-2 (1-686) x US-09-432-473-1 (1-5145)
Qy 11 CysGlySerValAlaThrProLeuGlyProLysTrpProGluProValPheGlyArgLeu 30
      |||||::: |||||
Db 2103 TGTGGTGGACTTCTTACCAAACTTAAC-----GGCACCATTA 2138
      |||||::: |||||
Qy 31 AlaSerProGlyPheProGlyGluTyrAlaAsnAspGluGluArgTrpThrLeuThr 50
      |||||::: |||||
Db 2139 ACCACCCCTGGCTGGCCCAAGAGTACCCCTTAATTAAGAAAGTGTGTGCAAGTGTT 2198
      |||||::: |||||
Qy 51 AlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuLeuSerHis 70
      |||||::: |||||
Db 2199 GCACCAACCACGATACAGAAATTTCTGTGAATTTGACTTTTAAATTTGGAAGCAATGAC 2258
      |||||::: |||||
Qy 71 LeuCysGluTyrAspPheValLys-----LeuSerSerGlyAlaLysValLeu 86
      |||||::: |||||
Db 2259 GTTTCGAAATATATATATGTGAGATCTGAGTGTCTTCTCTGAGTCAATTAATGCAAT 2318
      |||||::: |||||
Qy 87 AlaThrLeuCysGlyGluGlnGluSerThrAspThrGluArgAlaProGlyLysAspThrPhe 106
      |||||::: |||||
Db 2319 GGCAAATTCGTGGCGCTGAAGTG-----CCTGAAGTGATC 2354
      |||||::: |||||
Qy 107 TyrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrSerAsnGlyLysPro 126
      |||||::: |||||
Db 2355 ACATCCAGTGTCAACATATAGCAATTTGAATTTCAATTCGACAAATGCTGATCCAAAG--- 2411
      |||||::: |||||
Qy 127 PheThrGlyPheGluAlaPheTyrAlaValGluAspIleAspGluCysGluValAlaPro 146
      |||||::: |||||
Db 2412 --AAGGCTTCAAACACACATTTTCTCA--GACAAAGATGAATGCTTAAGATTAAT 2465
      |||||::: |||||
Qy 147 GlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGlyLysPheTyrCysSer 166
      |||||::: |||||
Db 2466 GGT-----GGATGTACGACGAGATGTGTCAACACAGATGGGAGTACATGTCTCAA 2516
      |||||::: |||||
Qy 167 CysArgAlaGlyTyrValIleuHisArgAsnLysArgThrCysSerAlaLeuCysSerGly 186
      |||||::: |||||
Db 2517 TGCCGTAATGGAATTTGTCTACTACATGACAAATTAACATGATTTGCAGAGAGCTGATGAA 2576
      |||||::: |||||
Qy 187 GlnValPheThrGlnArgSerGlyLysLeuSerSerProGluTyrProArgProTyrPro 206
      |||||::: |||||
Db 2577 CAGAAATCCACAGTCCCAAGTGGCTTCATCACCAAGTCCCAACAGTCCAGACAGATACCA 2636
      |||||::: |||||
Qy 207 LysLeuSerSerCysThrTyrSerIleSerLeuGluGlnGlyPheSerValIleuAsp 226
      |||||::: |||||
Db 2637 AGCAGGAAGAATGCACTGGAAATCAGCCGACCTCTGGCCACCGAATCAATGAGCC 2696
      |||||::: |||||
Qy 227 PheValGluSerPheAspValGluThrHisProGluThrLeuCysProTyrAspPheLeu 246
      |||||::: |||||
Db 2697 TTATAGTAA--TTTGAGATTGAGCAGCATCAAGAA--TGCTCTATGACCACTTA 2747
      |||||::: |||||
Qy 247 LysIleGluThrAspArgGluGluHisGlyPro-----PheCysGlyLysThr 262
      |||||::: |||||
Db 2748 GAAGTATTGATGGAGAAACAGAAAGTCAACCGATTTCTTGAGAGACTATGTGCAACAG 2807
      |||||::: |||||

```



```
Qy 207 LysleuSerSerCysThrTyrSerIleSerLeuGluGlyPheSerValIleLeuasp 226
Db 1182 AGCAGAAAGAAATGCATCTGGGAATCAGCGCCACTCTCGGCACCGAATCAAAATTAAGC 1241
Qy 227 PheValGluSerPheAspValGluThrHisProGluThrLeuCysProTyrAspPheLeu 246
Db 1242 TTTTACTGAA---TTTGAGATTGAGCAGCATCGGAA---TGTGCTTATGACCACCTTA 1292
Qy 247 LysIleGluThrAspArgGluGluHisGlyPro-----PheCysGlyLysThr 262
Db 1293 GAAAGATTGATGAGAAACAGAAAGATCACCGATTCTTGACAGACTATGTGGCAACAG 1352
Qy 263 LeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGlu 282
Db 1353 ATACAGATCCCTCTGTGGCTACTGGAAATAAATGTTGTTCGTTGTTCTTCATGCA 1412
Qy 283 SerGlyAspHisThrGlyTyrPheLysIleHisTyrThrSer-----Thr 296
Db 413 TCTGTTCAGAAAGAAAGCGCTTTCACACCATCTCTACAGAGTGTGGCGAGATTGAAA 1472
Qy 297 AlaGlnProCysPro-----TyrProMetAlaProProAsnGlyHisVal 311
Db 1473 GCAGATCAAAACAGAGATCTGTACTCACATGCTCAGTTGGTATACAACTAACCCA 1532
Qy 312 SerProValGlnAlaLysTyrIleLeuLysAsp-----Ser 323
Db 1533 GGACAGGTGACTGTGATGGCTATTATATCAGAAAGGGGCTCTCCACTTCATTAATCC 1592
Qy 324 PheSerIlePhe-----CysGluThrGlyTyrGluLeuGluGlnGlyHis 338
Db 1593 TTCCAGACATTTGAGAGTGAGAGAGACAGACAGCTGGCTATGACTATGACTGACCTCTTT 1652
Qy 339 LeuProLeuLysSerPheThrAlaVal-----CysGlnLysAspGlySer 353
Db 1653 GATGCTCTTGATCA---ACAGCTGTGGGCTTGCTGCATCTGT-----GGATCC 1700
Qy 354 TrpAspArgProMetProAlaCysSerIleValAspCysGlyProProAspAspLeuPro 373
Db 1701 -----GGGCCACCGAAGAGATTAT 1721
Qy 374 Ser----- 374
Db 1722 TCATTTGAGATTCAGTTTAAATTCATTTCCACACTGATGACACAAATCAAGAAAGGA 1781
Qy 375 -----GlyArgValGluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIle 392
Db 782 TTTCTATTAAGTACAAAGCATTAAGATATCCAGATACCACATACCAAAAATTAACAC 1841
Qy 393 GlnTyrSerCysGluGluThr-----PheTyr 401
Db 1842 CAAAACCTCTGTACAAACAAAGAAATGTGCATATATGAGAGAAACATATTTTTTTA 1901
Qy 402 ThrMetLysValAsnAspGlyLysTyrValCysGluAlaAspGlyPheThrThrSerSer 421
Db 1902 AAACGGAAGATA-----TTGGCACAAAATTTTATCAAAAGAGT 1940
Qy 422 LysGlyGluLysSerLeu 427
Db 1941 TTGAACAAAATCCCTG 1958

RESULT 20
US-09-432-473-3
: Sequence 3, Application US/09432473
: Patent No. 6365715
: GENERAL INFORMATION:
: APPLICANT: ARLETH, ANTHONY J.
: APPLICANT: WILLETTTE, ROBERT N.
: APPLICANT: ELISHOURBAGY, NABILI A.
: TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLUOID-LIKE PROTEIN
: FILE REFERENCE: ATG-50038-D1
: CURRENT APPLICATION NUMBER: US/09/432,473
```

```
: CURRENT FILING DATE: 1999-11-01
: EARLIER APPLICATION NUMBER: 08/991,408
: EARLIER FILING DATE: 1997-12-16
: EARLIER APPLICATION NUMBER: 60/034,471
: EARLIER FILING DATE: 1997-01-02
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 3690
: TYPE: DNA
: ORGANISM: HOMO SAPIENS
: US-09-432-473-3

Alignment Scores:
Pred. No.: 6,79e-32 Length: 3690
Score: 410.00 Matches: 133
Percent Similarity: 41.15% Conservative: 67
Best Local Similarity: 27.37% Mismatches: 168
Query Match: 10.91% Indels: 118
DB: Gaps: 20

US-09-874-198-2 (1-686) x US-09-432-473-3 (1-3690)
Qy 11 CysGlySerValAlaThrProLeuGlyProLysTyrProGluProValPheGlyArgLeu 30
Db 648 TGTGTGAGACTTCTTACCAACTTAAC-----CGCACCATTA 683
Qy 31 AlaSerProGlyPheProGlyLysTyrAlaAsnAspGlnGluArgTyrThrLeuThr 50
Db 684 ACCACCCCTGTGGCCCAAGAGATCCCTCTTAATAGAACCTGTGTGGCAACTGTGTT 743
Qy 51 AlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuGluLeuSerHis 70
Db 744 GCACCAACCACTACAGATTTCTGTGAAGTTTGACTTTTGAATTTGGAAGCAATGAA 803
Qy 71 LeuCysGluTyrAspPheValLys-----LeuSerSerGlyAlaLysValLeu 86
Db 804 GTTTCAAATATGATTAATGTGAGATCTGAGTGGCTTCTTCTGTGAGCTTAACCTCAT 863
Qy 87 AlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAlaProGlyLysAspThrPhe 106
Db 864 GGCAAAATCTGTGGCGCTGAAGTG-----CCTGAAGTGATC 899
Qy 107 TyrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrSerAsnGluLysPro 126
Db 900 ACATCCCACTTAACAATATGAGATTGATTAATCAATCTGACAAATACCTGATCCAAAG--- 956
Qy 127 PheThrGlyPheGluAlaPheTyrAlaIleGluAspIleAspGluCysGlnValAlaPro 146
Db 957 ---AAGGCTTCAAGACATTTTTCCTCA---GACAAAGATGAATGCTCAAGATATAT 1010
Qy 147 GlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGlyGlyPheTyrCysSer 166
Db 1011 GGT-----GGATGTCAGACGAAATGTCTCAACAGATGCGAGCTACCTATGCTCAA 1061
Qy 167 CysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSerGly 186
Db 1062 TCCCGTAATGATTTGCTCTACATCAATTAACATGATTCGACAAAGAACTAGTGTAA 1121
Qy 187 GlnValPheThrGlnArgSerGlyLysLeuSerSerProGluTyrProArgProTyrPro 206
Db 1122 CAGAGATCCACAGTCCAAAGTGGCTCATACACAGTCCCAACTGGCCAGACAAATGACCA 1181
Qy 207 LysleuSerSerCysThrTyrSerIleSerLeuGluGluGlyPheSerValIleLeuasp 226
Db 1182 AGCAGAAAGAAATGCATCTGGGAATCAGCGCCACTCTCGGCACCGAATCAAAATTAAGC 1241
Qy 227 PheValGluSerPheAspValGluThrHisProGluThrLeuCysProTyrAspPheLeu 246
Db 1242 TTTTACTGAA---TTTGAGATTGAGCAGCATCGGAA---TGTGCTTATGACCACCTTA 1292
Qy 247 LysIleGluThrAspArgGluGluHisGlyPro-----PheCysGlyLysThr 262
Db -----
```

```

Db 1293 GAAGTATTGATGAGAAAACAGAAAAGTCACCGATTCTTGACGACTATGTGCAACAAG 1352
QY 263 LeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGlu 282
Db 1353 ATACCAGATCCCCCTGGCTACTGGAATTAATGTTGGTTGGTTGTTTCTATGCA 1412
QY 283 SerGlyAspHisThrGlyTrpLysIleHisThrSer-----Thr 296
Db 1413 TCTGTCAAGAAAAGCGTTTCAACCCACACATCTACAGACTGTGGCGGAGATTGAAA 1472
QY 297 AlaGlnProCysPro-----TyrProMetAlaProProAsnGlyHisVal 311
Db 1473 GCAGATCAAAACCAAGAGATCTACTACATGCTCAGTTGGTGAATCAACATACCA 1532
QY 312 SerProValGlnAlaLysTrpIleLeuLysAsp-----Ser 323
Db 1533 GCACAGGTTGACTGTGAATGGCTATTAGTATGACAAAGGGGCTTCGACTTAATTATCC 1592
QY 324 PheSerIlePhe-----CysGluThrGlyTyrGluLeuLeuGlnGlyHis 338
Db 1593 TTCCAGACATTTGAAGTGGAGAGACAGACAGACTGTGGCTATGACTATGTGAGCTCTT 1652
QY 339 LeuProLeuLysSerPheThrAlaVal-----CysGlnLysAspGlySer 353
Db 1653 GATGCTCTTGATTCA--ACAGCTGTGGGGCTTGGTCGATTCTGT-----GGATCC 1700
QY 354 TrpAspArgProMetProAlaCysSerIleValAspCysGlyProProAspAspLeuPro 373
Db 1701 -----GGCCACACAGAAAGATTAT 1721
QY 374 Ser----- 374
Db 1722 TCAATTGGAGATTCAAGTTTATTCATTTCACACTGATGACACATCAACAAGAAGGA 1781
QY 375 -----GlyArgValGluTrpIleThrGlyProGlyValThrThrLysAlaValIle 392
Db 1782 TTTCATATATAGTACAAAAGCATAGATATCCAGATACACACATACCAAAAATAACAC 1841
QY 393 GlnTyrSerCysGlnGluThr-----PheTyr 401
Db 1842 CAAAACCTCTGTCAACACAAAGAAATGTCATATATGAGAGAAAGACATATTTTATTA 1901
QY 402 ThrMetLysValAsnAspGlyLysTyrValCysGlnAlaAspGlyPheTrpHisSer 421
Db 1902 AACTGGAAGAT-----TTGGCACAATGTTTATACAAAGAT 1940
QY 422 LysGlyGluLysSerLeu 427
Db 1941 TTGAACAAAATAATCCCTG 1958
QY 21
Db 1958-200-900A-1
Sequence 1, Application US/08200900A
Patent No. 5665566
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc. - Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,900A
FILING DATE: 23-FEB-1994

```

```

: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Meindert Maureen C.
: REGISTRATION NUMBER: 31,544
: REFERENCE/DOCKET NUMBER: GI 5201-FWC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 876-1170 X8574
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2581 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-200-900A-1

Alignment Scores:
Pred. No.: 5,42e-32 Length: 2581
Score: 408.50 Matches: 150
Percent Similarity: 42.39% Conservative: 95
Best Local Similarity: 25.95% Mismatches: 193
Query Match: 10.87% Indels: 141
DB: 1 Gaps: 31

US-09-874-198-2 (1-686) x US-08-200-900A-1 (1-2581)
QY 172 ValLeuHisArgAsn---LysArgThrCysSerAlaLeu-----CysSerGlyGlnVal 188
Db 882 CTCCTCCACACAGAACTCCACAGGAGCTGAGGGCTCATGACCTGAGGCCAAATA 941
QY 169 PheThrGlnArgSerGlyGluLeuSerSerProGluTyrProArgProTyrProLysLeu 208
Db 942 CA-ACA-----TTCAAGTCTATAAATTCCCAACAGCTTACCTTATATGAG 985
QY 209 SerSerCysThrTyrSerIleSerLeuGlnGlyPheSerValIleLeuAspPheVal 228
Db 986 GCTTCTGATTTGGAATTTAATGACAAAAGGAAAAAATATACACTCCACTTCA 1045
QY 229 GluSerPheAspValGluThrHisProGluThrLeuCysProTyrAsp----- 244
Db 1046 GAA---TTTGACCTGGAAAATATTCAGATGATGTAATTCAGAAATCAGAGTGGAAAGAGAT 1102
QY 245 -----PheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGlyLys 261
Db 1103 GATTCCTGTTTATGAGTGTACACAGC-----CTGGTCCGTA----- 1144
QY 262 ThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAsp 281
Db 1145 -----AACGATGTGTTCACACCAACCAAGATGCTGCTTTATACATGAT 1195
QY 282 GluSerGlyAspHisThrGlyTrpLysIleHisThrSer-----Thr 296
Db 1196 AATATGCTGGCAAAACAGGAGATTAAACCAAAATTTCACTAGCTAGGTAGCTTGGGAGATT 1255
QY 297 AlaGlnProCysProTyrProMetAlaProProAsnGlyHisValSerProValGlnAla 316
Db 1256 CCAGAACCTGCACAG-----TCACAGATGGG----- 1270
QY 317 LysTyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGln 336
Db 1271 -----GAAGACAATTTTCAG-----TCCAGAGATGGG----- 1297
QY 337 GlyHisLeuProLeuLysSerPheThrAlaValCysGln-----LysAsp 351
Db 1298 GAGTGTATCCGCTGTGTAAT-----CTCTGTGAGGTTTCCACACTGTAGAGAT 1348
QY 352 GlySerTrpAspArgProMetProAlaCysSerIleValAspCysGlyProProAspAsp 371
Db 1349 GECTCAGATGAA-----GCACACTGTGTGCTCTTCAATGAGCAGACAGACAGC 1399
QY 372 LeuProSerGlyArgValGluTrpIleThrGlyProGlyValThrThrTyrLysAlaVal 391

```



```

Db 1349 GGCCTCAGATGAA-----GCACACTGTGTGCTCTTCAATGCACAGACAGACACC 1399
QY 372 LeuProSerGlyArgValGlyIleThrGlyProGlyValThrThrIleVal 391
Db 1400 -----AGTGGTTGGTGGAGTTTC-----AGATCCAAAGCATA 1432
QY 392 IleGlnTrpSerCysGluGluThrPheThrMetLysValAsnAspGlyLysTyrVal 411
Db 1433 TGGCAGTACCTCTGGCCGAGAAC---TGGACAACCAAGATCTCAGATGAT-----GTG 1483
QY 412 CysGluAlaAspGlyPheThrPheSerSerLysGlyLysSerLeuProValCysGlu 431
Db 1484 TGTCACTGCTGGAGTGGAGGACT-----GGAAACTATCCGCGCCACACTTTTCT 1534
QY 432 -----ProValCysGlyLeuSerAlaArgThrGlyGly----- 443
Db 1535 ACTGAGAGTGGACCATGTGTAATTTAAACACAGCACCTTAATGACCTTAATACTAAGC 1594
QY 443 ----- 443
Db 1595 CCAAGCCACAGTGTAGAGATTCACGTATTTGCTACATGTAACATAATCATGT 1654
QY 444 -----ArgLleTyrGlyGlyLysAlaLys 452
Db 1655 GGGAAAAAACTGGTGAAGTCAAGAAAGTTAGCCCGAAGATGTCCGAGAGAGTCCACGA 1714
QY 453 ProGlyAspPheProThrPheVal-----LeuLleuGlyLysThrThrAlaAlaGlyAla 471
Db 1715 GAAAGAGCCCTGGCCCTGGCTGTCTGTATTTGACAGCATCAACAGTCTCGGGAGCT 1774
QY 472 ---LeuLeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysHis 490
Db 1775 TCTCTGGTGAAGAGGATGTGCTGTGTGGCCGCCACGCGGTGlnAGGGAGAAATATG 1834
QY 491 AspAlaSer-----AlaLeuAspLleArgMetLysThrLeuLysArgLeuSer 506
Db 1835 GACCCGCTAGTGAAGAACGATGCTAGCCCATGTGACATCA---AATCTGACTTCT 1891
QY 507 ProHisThrGlnAlaThrSerGlnAlaValPheLleHisGluGlyTyrThrHisAsp 526
Db 1892 CCTCAGATGAACACTAGTGTGATGACCAATTTGCTTAACCCACACACTCA---AATAAA 1948
QY 527 AlaGlyPheAspAsnAspLleAlaLeuLysLeuAsnAsnLysValValLleAsnSer 546
Db 1949 CGGAGAAAGAACATGATGATGCCATGATGCATCTGAATGAAGTGAAGTCAACAGAT 2008
QY 547 AsnLleThrProLleCysLeuProArgLysGlnAlaGluSerPheMetArgThrAspAsp 566
Db 2009 TATATACAGCTATTTGTTTACCA-----GAAGAAAATCAAGTTTTTCCCGCAGGAAGA 2062
QY 567 IleGlyThrAlaSerGlyTyrGly---LeuThrGlnArgLysPheLeuAlaArgAsnLeu 585
Db 2063 ATTTGTTCTTGTGCTGTGGCTGGGGGACCTTATATATCAAGGTTCTACTGACAGACTG 2122
QY 586 MetTyrValAspLleProLleValAspHisGlnLysCysThrAlaAlaTyrGluLysPro 605
Db 2123 CAAAGAGCTAGCTCCCTCTTATCAATGAGAAATGT-----CAACAA 2167
QY 606 ProTyrProArgLysValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGly 625
Db 2168 CAGATGCCAATAATATACATTACGAAATAATGTGTGTCTCAGGCTATGAGACAGAGGG 2227
QY 626 LysAspSerCysArgLysAspSerGlyAlaLeuValPheLeuAspSerGluThrGlu 645
Db 2228 GTAGATTCTTGTCAAGGGGATTCAGGGGACACTC-----ATGCCCAAGAAAAACAC 2281
QY 646 ArgTrpPheValGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 665
Db 2282 AGATGGCTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 2338
QY 666 TyrGlyValThrLysValLleAsnTyrLleProThrLleGluAsnLleLle 683

```

```

Db 2339 CCAAGGGGTATGATCCCGGGTCCCAAGTTCCACAGAGTGGATACAAAGTTTCTTA 2392
RESULT 23
US-09-202-101-16
Sequence 16, Application US/09202101
Patent No. 6277618
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Recombinant blood-coagulation proteases
NUMBER OF SEQUENCES: 17
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/202,101
FILING DATE:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1389 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-202-101-16
Alignment Scores:
Pred. No.: 4,54e-32 Length: 1389
Score: 405.00 Matches: 107
Percent Similarity: 44.73% Conservative: 50
Best Local Similarity: 30.48% Mismatches: 92
Query Match: 10.78% Indels: 102
Gaps: 16
US-09-874-198-2 (1-686) x US-09-202-101-16 (1-1389)
QY 396 CysGluGluThrPheThrMetLysValAsnAspGlyLysTyrValCysGluAlaAsp 415
Db 421 TGGCAGACAG-----TTTGTAAATAATAGTCTGATTAACAGAGTGGTCTGCTGTACT 474
QY 416 GlyPheThrPheSerSerLysGlyLysSerLeuProValCysGluProVal----- 433
Db 475 GAGGATATCGACTTCAGAAACACAGAAAGTCC-----TGTGAACACAGACGTGCCA 525
QY 434 ---CysGly---LeuSerAlaArgThrThrGly----- 442
Db 526 TTTCCATGTGGAAGAGTTCTGTTCACAAACTTTTAAGCTACCCGCTGAGACTGTT 585
QY 442 ----- 442
Db 586 TTTCCGATGTGACTATGTAAATTCTACTGAGAGTGAACCATTTTGGATACATCACT 645
QY 443 -----GlyArgLleTyrGlyGlyLysAlaLysPro 453
Db 646 CAAAGACCCCAATCATTTATGACTTCACCTCGGGTGTGGTGGAGAAAGTCCAAACCA 705
QY 454 GlyAspPheProThrGlnValLeuLleLeuGlyGlyThrThrAla-----AlaGlyAla 471
Db 706 GGTCAATTCCTGGACAGGTGTGTGAATGTGAAGTTGATGATCATTTCTGTGGAGGCTCT 765
QY 472 LeuLeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaVal----- 485
Db 766 ATGCTTAATGAATAATGAGATTTGAACCTGCTGCCCACTGTGTGAACCTGTTAAATT 825
QY 486 -----TyrGluGlnLysHisAsp 491
Db 826 ACAGTTGTGCGAGGTGAACATAATATTGAGAGACAGAACATACAGACAAAAGCAAAAT 885
QY 492 AlaSerAlaLeuAspLleArgMetGlyThrLeuLysArgLeuSerProHisTyrThrGln 511
Db 886 -----GTGATTCGAATTAATTCCTACCAACAAC--- 912

```



```

1  NUMBER OF SEQUENCES: 15
2  CORRESPONDENCE ADDRESSES:
3  ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
4  ADDRESSEE: Cecchi, Stewart & Olstein
5  STREET: 6 Becker Farm Road
6  CITY: Roseland
7  STATE: New Jersey
8  COUNTRY: USA
9  ZIP: 07068
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: 3.5 inch diskette
12 COMPUTER: IBM PS/2
13 OPERATING SYSTEM: MS-DOS
14 SOFTWARE: Wordperfect 5.1
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/106.078
17 FILING DATE: 13-AUG-1993
18 CLASSIFICATION: 424
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER:
21 FILING DATE:
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Olstein, Elliot M.
24 REGISTRATION NUMBER: 24,025
25 REFERENCE/DOCKET NUMBER: 271010-160
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 201-994-1700
28 TELEFAX: 201-994-1744
29 INFORMATION FOR SEQ ID NO: 15:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 1548 bases
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: cDNA
36 FEATURE:
37 NAME/KEY: Factor IX cDNA
38 US-08-106-078-15
39
40 Alignment Scores:
41 Pred. No.: 5.42e-32 Length: 1548
42 Score: 405.00 Matches: 107
43 Percent Similarity: 44.73% Conservative: 50
44 Best Local Similarity: 30.48% Mismatches: 92
45 Query Match: 10.78% Indels: 102
46 DB: 1 Gaps: 16
47
48 US-08-874-198-2 (1-686) x US-08-106-078-15 (1-1548)
49
50 396 CysGslgluThrPheThrMetLysValAsnAspGlyLysTyrValCysGluValAsp 415
51 |||||:||||| ||| ||| |||||
52 426 TCGGAGCAG-----TTTGTAAATAATAGTCGTGATAAACAAGGGGTTTCTCTCTACT 479
53
54 416 GlyPheTrpPheSerLysGlyGluLysSerLeuProValCysGluProVal----- 433
55 ::|||:||||| |||||
56 480 GAGGATATGCACCTTGCAGAAACACGAAAGTCC-----TGTGAACACAGACAGTGCCA 530
57
58 434 -----CysGly-----LeuSerAlaIarGthrGrly----- 442
59 |||||:||||| |||
60 531 TTTCATGTGGGAAGAGTTGTGTTCACAAACTTCAAGCTACCCGCTGAGACGTGT 590
61
62 442 ----- 442
63
64 591 TTTCCTGATGTGACATATGTAAATTCTACGTGACGTGAACCATTTTGGATTAACATCAT 650
65
66 443 -----GlyArgIleTyrGlyGlyGluLysAlaLysPro 453
67 |||||:||||| |||||
68 651 CAAGACACCATCATCTTAATGACTTCACCTCGGGTGTGGTGGAAGAAGATGCCAAACCA 710
69
70 454 GlyAspPheProTrpGluValLeuIleLeuLysGlyThrThrAla-----AlaGlyAla 471
71 ||||| ||||| |||||
72 711 GGTCAATTCCTCTGGCAGAGTTGTGTGATGTAAAGTTGATGATTCATTCGTGTGAGGCTCT 770

```

Oy	472	LeuLeuTyrAspSerPheValLeuThrAlaAlaVal-----	485
Db	771	ATCGTATATGAAATATGATATTGTAAGTCTGCCCACTGTGTGAACATCGTGTAAATT	830
Oy	486	-----TyrGluGlnHisAsp	491
Db	831	ACAGTTGTCGAGTGAACATATATTTAGAGACACAGAACATATACAGCAAAAGGAAT	890
Oy	492	AlaSerAlaLeuAspIleArgMetGlyThrIleuLysArgLeuSerProHisIleTyrGln	511
Db	891	-----GTCATTGCAATTATTCTTCACACACAC---	917
Oy	512	AlaTyrSerGluAlaValPheIleHisGluGlyTyrThrHisAspAlaGlyPheAspAsn	531
Db	918	---TACATGACGCTATT-----AATTAAGTACAAACCA-----	947
Oy	532	AspIleAlaLeuIleLysIleAsnAsnLysValValIleAsnSerAsnIleThrProIle	551
Db	948	GACATATGCCCTCTCGAAGACGACAAACCTTATGTCCTAAACAGCTACCTATACACTATT	1007
Oy	552	CysLeuProArgLysGluAlaGluSerPheMetArgThrAspAspIleGlyThrAlaSer	571
Db	1008	TGCATTGCTGACAAAGAAATACAGAACATCTTCCATAATTGGATCTGGCATGTAGT	1067
Oy	572	GlyTyrPgly---LeuThrGlnArgGlyPheLeuAlaArgAsnLeuMetGlyValAspIle	590
Db	1068	GGCTGGGGGAAGATCTTCCAAAGAGAGATCAGCTTATGTTCTTCAGTACCTTAGACTT	1127
Oy	591	ProIleValAspHisGlnLysCysThrAlaAlaTyrGluLysProProTyrProArgGly	610
Db	1128	CCACTGTGTGACCGACGACATGTCTTCATCT-----ACAAGATTC	1169
Oy	611	SerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGlyLysAspSerCysArg	630
Db	1170	ACCATCTATTACAAACATGTTCTGCTGGCTTCCATGAAGAGGTAGATTCATGCTAA	1229
Oy	631	GlyAspSerGlyGlyAlaLeuValPheLeuAspSerGlyThrGlu---ArgTyrPheVal	649
Db	1230	GGAGTATAGTGGGGACCCCATGTT-----ACTGAAGTGGAAAGGACCAAGTTTCTTA	1280
Oy	650	GlyGlyIleValSerTyrPglySerMetAsnCysGlyGluAlaGlyGlnTyrGlyValTyr	669
Db	1281	ACGTGAATTATTACTGGGGTGA---GAGTGTGCATGAAGCAAAATATGCAATATAT	1337
Oy	670	ThrLysValIleAsnTyrIleProTyrPheGlu	680
Db	1338	ACCAAGGATTCGCGTATGTCACTAGGATTAAAG	1370

Search completed: January 24, 2003, 00:19:29
Job time : 148 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 23, 2003, 22:24:17 ; Search time 2217 Seconds
(without alignments)
5011.323 Million cell updates/sec

Title: US-09-874-198-2
Perfect score: 3758
Sequence: 1 MRLTLGLGSGVATPLGP.....GYTRKINVIPIENIISDF 686

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 75 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1p
-Q/cgpn2.1/USPFO-spool/US09874196/runat.22012003.085001.10895/app.query.fasta.1.839
-DB=EST -QEXT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=75
-DOCAALIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874196 @CGN.1.1.1716 @runat.22012003.085001.10895 -NCPU=6 -ICPU=3
-NO_XIPPY -NO_MAP -LARGEOURV -NEG.SCORES=0 -WAIT -LONLOG -DEV.TIMEOUT=120
-NARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hc:*
9: gb_estcl:*
10: gb_estcl2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estclun:*
16: em_estclun:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fam:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1176.5	31.3	843	13	B1146955
2	1099	29.2	775	13	B1219788
3	1009.5	26.9	820	9	A1266911
4	1004	26.7	699	10	BB653589
5	976	25.0	575	9	A1248851
6	952	25.3	544	10	AA72485493
7	930	24.7	729	9	A1226038
8	914	24.3	739	13	B6916805
9	878	23.4	600	12	BF532409
10	878	23.4	884	12	BF236712
11	833.5	22.2	678	12	BF233993
12	827	22.0	475	9	AA706982
13	808	21.5	752	12	BF233997
14	785.5	20.9	548	9	AL022666
15	777	20.7	586	9	A1529228
16	775	20.6	520	9	AA337253
17	767	20.4	620	14	BM780799
18	766.5	20.4	868	13	BG921224
19	758	20.2	552	9	AA238260
20	741	19.7	939	14	B0644570
21	738	19.6	1035	14	B0653181
22	736	19.6	913	14	B0648641
23	736	19.6	921	14	B0648719
24	736	19.6	925	14	B0645367
25	735	19.6	796	14	B0650766
26	735	19.6	911	14	B0654231
27	723	19.5	523	9	A1327510
28	723	19.2	106	12	BF159541
29	698	18.6	598	9	A1391397
30	689	18.3	700	13	BG921962
31	676.5	18.0	472	9	A1099288
32	659.5	17.8	487	9	A1048966
33	647.5	17.5	550	13	BF1715538
34	647.5	17.2	922	9	AL548397
35	642	17.1	977	9	AL557931
36	632.5	16.8	384	13	B1789626
37	631	16.8	480	12	BF6840188
38	627	16.7	915	14	B0939021
39	621.5	16.5	473	9	AA764061
40	619	16.5	447	14	R56666
41	619	16.5	452	13	B1847411
42	614	16.3	734	14	B0771154
43	610.5	16.2	832	13	B1825421
44	607.5	16.2	669	12	BE890174
45	605	16.1	468	9	AA945177
46	604	16.1	731	14	BM947443
47	603	16.0	698	10	BB623493
48	602	16.0	861	13	BM008361
49	602	16.0	915	14	B0898913
50	599.5	16.0	432	10	BB800289
51	598	15.9	940	9	AL541720
52	593.5	15.8	905	9	AL540482
53	591.5	15.7	991	14	B0723665
54	590	15.7	416	14	W14100
55	587.5	15.6	894	9	AL540720
56	587	15.6	881	9	B0881801
57	583	15.5	450	9	AA208933
58	582.5	15.5	903	14	B0893389
59	575	15.3	887	14	B0880735
60	575	15.3	939	14	B0894117
61	571.5	15.2	790	13	B1330963
62	564.5	15.0	1051	14	B0067932
63	564	15.0	802	13	B1454262
64	562.5	15.0	816	12	BG685001

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM1239 row: e column: 06
 High quality sequence stop: 775.
 Location/Qualifiers

FEATURES

source

1. 775
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5099501"
 /clone_lib="NCI_CGAP_L19"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP library."

BASE COUNT

182 a 235 c 195 g 163 t

Alignment Scores:

Score: 2.05e-105 Length: 775
 Percent Similarity: 1099.00 Matches: 208
 Best Local Similarity: 88.28% Conservative: 18
 Query Match: 81.25% Mismatches: 28
 DB: 29.24% Indels: 4
 Gaps: 1

US-09-874-198-2 (1-686) x B1219788 (1-775)

Qy 61 PherhrlhspheaspheugluleuSerhslsleuCygluTyraSpheValylsleuSer 80
 Db 2 TTCACCCACTTGTACCTGGAACCTCTTACCGCTGCGAGTATGATCTTGTCAAGTTGAGC 61
 Qy 81 SerGlyAlaValleuAlaThrleuCygluGlnGlnSerThrAspThrGluAla 100
 Db 62 TCAGGACCAAGGTGCTGCCACACTGTGTGGCAGAGATACAGACACTGACGAGCA 121
 Qy 101 ProGlyLysAspThrPheTySerLeuGlySerLeuAspThrPheAspSer 120
 Db 122 CCGGCAATGACACCTTCACTACAGGGTCCACGCTTAAGGTCACTTCCACTCGAC 181
 Qy 121 TyrSerAsnGlnLysProPheThrGlyPheGluAlaPheTyraAlaGluAsp 140
 Db 182 TACTCCAAATGACAAGCCGTTCAACAGGCTTGTGAGGCTTATGACACGAGATGTGAT 241
 Qy 141 GlucyGlnAlaValAlaProGlyLysAlaProThrCysAspHisCysHisAsnHisLeu 160
 Db 242 GAATGACAGGTGCTCTGGGAGACTGATCCCTTGTGACCAATTATTCACCAACTACTTG 301
 Qy 161 GlyGlyPheTyCysSerCysArgAlaGlyTyraValleuHisArgAsnLysArgThrCys 180
 Db 302 GCGGCTATATTGCTCCCGCAGAGGGGCTACGTTCTCCACAGACAAGCACACACTGCG 361
 Qy 181 SerAlaLeuCySerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGlu 200
 Db 362 TCAGCCCTTGTTCAGGGCAGGCTTCCACAGAAATCTGGGATCTCAGTACCCCTGAG 421
 Qy 201 TyrProArgProTyProLysLeuSerCysThrTySerTleSerLeuGlnGlnGly 220
 Db 422 TACCCACACACATACCCCAAGCTTCCACCTCAGCTACAGCATCGGCTGGAGGAGCGC 481
 Qy 221 PheSerValAlaLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeu 240
 Db 482 TTCAGTGTATCTGGAGCTTCTGGAGTCTTCGATGTGAGAGACGACCTTGAAGCCGAG 541
 Qy 241 CysProTyraSpheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGly 260
 Db 542 TCACCTATGACTCCCTCAAGATTCACAAGACAGAGGAGGAGACAGGCCCATTTGTGGG 601
 Qy 261 LysThrLeuProHisArgIleGlnThrLysSerAsnThrValThrIleThrPheValThr 280
 Db 602 AAGACCTCTCTCCACGATTAAGTGAACCTGACAGCCACAGGAGTACCATCTTGTCCACT 661

Qy 281 AspGluSerGlyAspHisThrGlyTrp-LysIleHisTyrrhrSerThrAlaGlnProcy 300
 Db 662 GACGACTCGGGGAGACACAGGCTGGAGAAATACACTACACAGACAGCAGCGGCC-16 720
 Qy 300 sProTyrrProMetAlaProProAsnGlyHisValSerProValGln 315
 Db 721 CCCTGATCCAAC-GGCCCACTATGACAGAT--TCACCTGTGACAG 762

RESULT 3

LOCUS AI266911/c 820 bp mRNA linear EST 16-NOV-1998
 DEFINITION uJ08e08.x1 Sugano mouse liver mla Mus musculus cDNA clone
 IMAGE:1891334 3' similar to FR:000187 000187 MASP-2 PROTEIN. ;,
 mRNA sequence.

ACCESSION AI266911
 VERSION AI266911.1 GI:3885069
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS
 Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 820)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE
 JOURNAL
 COMMENT
 The WashU-HIMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HIMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:975658

FEATURES
 source
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 503.
 Location/Qualifiers

1. 820
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1891334"
 /clone_lib="Sugano mouse liver mla"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: Liver; Vector: pME185-FL3; Site: 1; DraIII
 (CACTGTGTG); Site: 2; DraIII (CACCAGTGTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGCGGCTTTTGTGTTTTTTTTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTTCGCTACTGG], digested
 and cloned into distinct DraIII sites of the pME185-FL3
 vector (5' site CACTGTGTG, 3' site CACCAGTGTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTTCCTCTTAAGAAGCTGCG and 3' end
 primer CGACCTCGAGCTCGAGCA." primer

BASE COUNT 194 a 214 c 200 g 212 t

Alignment Scores:

Pred. No.: 6.57e-96 Length: 820
 Score: 1009.50 Matches: 215
 Percent Similarity: 42.10% Conservative: 22
 Best Local Similarity: 38.19% Mismatches: 32

Query Match: 26.86% Indels: 294
 DB: 9 Gaps: 2
 US-09-874-198-2 (1-686) x AI266911 (1-820)

QY 125 LysProPheThrGlyPheGluAlaPheTyrAlaAlaGluAspIleAspGluCysGlnVal 144
 |||||.....
 DB 820 AAGCCTTACAGAGGTTTACGCTTCTATGACGCGAGAGATGTGATGATCAATCCAAAGTG 761
 QY 145 AlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGlyGlyPheTyr 164
 ::|||:::|||||.....
 DB 760 TCTCTGGAGACTCCTCCTTGTGTGACATTATGCCCCACACTACGGCGGGCTTCTAT 701
 QY 165 CysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCys 184
 |||||.....
 DB 700 TCTCTCTGACAGAGCGGGTACGTTTCCACACAGAACAGACACGTCACAGCCCTTGT 641
 QY 185 SerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGluTyrProAlaGPro 204
 |||||.....
 DB 640 TCAGGCCAGGTGTTCACAGCAAGATCTGGGTATCTCAGTACGCCGTACCCACAGCCA 581
 QY 205 TyrProLysLeuSerSerCysThrTyrSerIleSerLeuGlnGlnGlyPheSerValIle 224
 |||||.....
 DB 580 TACCCCAAGCTCTCCAGCTGCACCTACAGCATCCGCTGAGAGCGGCTTCAAGTGCATC 521
 QY 225 LeuAspPheValGluSerPheAspValGluThrHisProGluThrLeuCysProTyrAsp 244
 |||||.....
 DB 520 CTGAGACTTCGTGGAGTCCCTTCGATGTGGAGACGACCTGAAGCCGACGCGCCCTATGAC 461
 QY 245 PheLeuLysIleGlnThrAspArgGluGluHisGlyProPheCysGlyLysThrLeuPro 264
 |||||.....
 DB 460 TCCCTTCAG----- 452
 QY 265 HisArgIleGluThrLysSerAsnThrValThrIleThrPheValThrAspGluSerGly 284
 452 ----- 452
 QY 285 ASPHisThrGlyTyrLysIleHisTyrThrSerThrAlaGlnProCysProTyrPromet 304
 452 ----- 452
 QY 305 AlaProProAsnGlnHisValSerProValGlnAlaLysTyrIleLeuLysAspSerPhe 324
 DB 452 ----- 452
 QY 325 SerIlePheCysGlnThrGlyTyrGluLeuLeuGlnGlyHisLeuProLeuLysSerPhe 344
 DB 452 ----- 452
 QY 345 ThrAlaValCysGlnLysAspGlySerTyrAspArgPrometProAlaCysSerIleVal 364
 452 ----- 452
 QY 365 AspCysGlyProProAspAspLeuProSerGlyArgValGluTyrIleThrGlyProGly 384
 DB 452 ----- 452
 QY 385 ValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMetLys 404
 DB 452 ----- 452
 QY 405 ValAsnAspGlyLysTyrValCysGluAlaAspGlyPheTyrThrSerSerLysGlyGlu 424
 DB 452 ----- 452
 QY 425 LysSerLeuProValCysGluProValCysGlyLeuSerAlaArgThrThrGlyGlyArg 444
 DB 452 ----- 452
 QY 445 IleTyrGlyGlyGlnLysAlaLysProGlyAspPheProTyrGlnValLeuIleLeuGly 464
 DB 452 ----- 452

QY 465 GlyThrThrAlaAlaGlyAlaLeuLeuTyrAspAsnTyrValLeuThrAlaAlaHisAla 484
 DB 452 ----- 452
 QY 485 ValTyrGluGlnLysHisAspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArg 504
 DB 452 ----- 452
 QY 505 LeuSerProHisTyrThrGlnAlaTyrSerGluAlaValPheIleHisGluGlyTyrThr 524
 DB 452 ----- 452
 QY 525 HisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValValIle 544
 DB 451 ----- 437
 QY 545 AsnSerAsnIleThrProIleCysLeuProAlaGlyGluAlaGluSerPheMetArgThr 564
 DB 436 AACGGAAGCATATGCTTGTGCTTACCGCGAAGAAAGAGCGCATCTTAATGAGAAACA 377
 QY 565 Asp-IleGlyThrAlaSerGlyTyrPglyLeuThrGlnArgGlyPheLeuAlaArgAs 584
 DB 376 GCCTTCCCTGGAACGTGCTGCTGGGGGTAAACCCAGAAAGGGCTTCTTGAGAAA 317
 QY 584 nLeuMetTyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGlyLys 604
 DB 316 CCAATGTTTGGACATPACCAATTGCTGACCAACCAAAATGTACCGCGCTGTAGAAA 257
 QY 604 sProProTyrProAlaGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerG 624
 DB 256 G---CTGTATCCAGGAGTAAAGTAAGCCGCTAACATGCTGTGCTGTAGAGACTGG 200
 QY 624 yGlyLysAspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluTh 644
 DB 199 TGGCAAGGACACTGTCAGAGGTGACAGTGGGGGCGATTAGTCTTCTAGATTAAGAGAC 140
 QY 644 rGluArgTyrPheValGlyGlyIleValSerTyrPglySerMetAsnCysGlyGluAlaG 664
 DB 139 ACAGCATGCTTGTGGGAGGAATAGTTCTCGGGTTCATTAATTTGGGGCGGCGACA 80
 QY 664 yGlnTyrGlyValTyrThrLysValIleAsnTyrIleProTyrPglyLeuAsnIleIle 684
 DB 79 CCAGTATGGGGTCTACACAAAGTCATCACTATTCCTCGATTGAGACATATAAG 20
 QY 684 rAspPhe 686
 DB 19 TAATTTC 13

RESULT 4
 BB653589 699 bp mRNA linear EST 26-OCT-2001
 LOCUS BB653589 RIKEN full-length enriched, adult male liver tumor Mus
 DEFINITION musculus CDNA clone C730007H22 5', mRNA sequence.
 ACCESSION BB653589
 VERSION BB653589.1 GI:16487417
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 699)
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Koude,
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyai,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 TITLE Unpublished (2001)
 JOURNAL Contact: Yoshitake Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute

[illegible][illegible]

US-09-874-198-2 (1-686) x BF532409 (1-600)

OY 1 MeArgLeuLeuThrLeuLeuGlyLeuLeuCysGlySerValAlaIthrProLeuGlyPro 20
|||||
Db 6 ATAGAGCTACTCTCTTCTGGGTCTGCTGTGAGATTGGTGGCCACACTTCTGGGTCA 65
OY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyTyrAla 40
|||||
Db 66 AAGTGCCTGAACCTGTATCTCGGGCCCTGTGCTCCCTGCTCCACAGAGATGCT 125

OY 41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
|||
Db 126 GACCATCAAGATGACCTGTGACACCTGACACCCCTGTGACCCGCTCTCTAC 185

OY 61 PheThrHisPheAspLeuGluLeuSerHisLeuCysGlyTyrAspPheValysLeuSer 80
|||||
Db 186 TTCACCCACTTTCACCTGGAACCTCTTACCGCTGCGAGTATACCTTTGCAAGTTGAGC 245

81 SerGlyAlaValLeuAlaThrLeuCysGlyGlnGluSerHisAspThrArgAlaGlyAla 100
|||||
246 TCAGGGACCAAGGTCTGGCCACACTGTGTGGCAGAGAGTACAGACACTGACAGGCA 305

OY 101 ProGlyLysAspThrPheThrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
|||||
Db 306 CCGGGCAATGACACCTTCTACTCTGCGGTCCAGCCTTAAGGTCACTCCACTCCGAC 365

OY 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheThrAlaIleGluAspIleAsp 140
|||||
Db 366 TACTCCAAATGAGAGCCGTTTCACAGGCTTTGAGGCTTCTATGACAGGGAGATGGAT 425

OY 141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisAsnHisLeu 160
|||||
Db 426 GAATGACAGAGTCTCTGGGAGACATCAGTCCCTGTGACCATTAATGCCACACTACTTG 485

OY 161 GlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
|||||
Db 486 GGGGCTACTATTGCTCCGACAGGGGCTGCTCTCCACAGAACAGCAGCAGCTGC 545

OY 181 SerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSer 198
|||||
Db 546 TCAGCCCTTTGTTCAAGCCAGGCTTTCACAGAGATCTGGGATCTCATGATGAC 599

RESULT 10
BF236712 884 bp mRNA linear EST 14-NOV-2000
LOCUS 602028006r1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4163448 5',
DEFINITION mRNA sequence.
ACCESSION BF236712 GI:11150598
VERSION BF236712.1 GI:11150598
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 884)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHM9447 row: k column: 01
High quality sequence stop: 615.
Location/Qualifiers
1..884
/organism="Mus musculus"
/strain="FVB/N"

/db_xref="taxon:10090"
/clone="IMAGE:4163448"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Liver; Vector: pcMW-SPOK6; Site: NotI;
Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 191 a 256 c 252 g 185 t

ORIGIN

Alignment Scores:
Pred. No.: 6,02e-82 Length: 884
Score: 878.00 Matches: 189
Percent Similarity: 79.77% Conservative: 20
Best Local Similarity: 72.14% Mismatches: 46
Query Match: 23.36% Indels: 10
DB: 12 Gaps: 3

US-09-874-198-2 (1-686) x BF236712 (1-884)

OY 7 LeuGlyLeuLeuCysGlySerValAlaIthrProLeuGlyProLysTrpProGluProVal 26
|||||
Db 6 CTGGCTGTGCTGTGAGATTG-GTGGCCACACTTCTGGGTCAAGTGGCCTGAACCTGTA 64

OY 27 PheGlyArgLeuAlaSerProGlyPheProGlyGlyTyrTrpAlaAsnAspGlnGluArg 46
|||||
Db 65 TTGGGCGCGCTGTGCTCCCTGCTCCACAGAGATGCTGACCATCAAGATGATGCC 124

OY 47 TrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeu 66
|||||
Db 125 TGACACATGACTACACCCCTGGCTACCGCTGCGCTGACTTCAACCCACTTGGACCTG 184

OY 67 GluLeuSerHisLeuCysGlyTyrAspPheValLysLeuSerSerGlyAlaValLeu 86
|||||
Db 185 GAACCTCTTACCGCTGCGAGATGATGACTTGTCAAGTTGAGCTCAGAGCAAGAGTGTG 244

OY 87 AlaThrLeuCysGlyGlnGluSerThrAspThrGlnArgAlaProGlyLysAspThrPhe 106
|||||
Db 245 GCCACACTGTGTGGGAGGAGTACAGACACTGACAGGCACTGGCAATGACACCTTC 304

OY 107 TyrSerLeuGlySerSerLeuAspIleThrPheArgSerAspTyrSerAsnGluLysPro 126
|||||
Db 305 TACTCACTGGGTCCAGCCCTTAAGTCCACTTTCACCTCGACATCTCCATGAGAACGCC 364

OY 127 PheThrGlyPheGluAlaPheThrAlaIleGluAspIleAspLysGlnValAlaPro 146
|||||
Db 365 TTCACAGGCTT-GAGGCTTCTATGACAGCGGAGAGATGTGATGAATGCAAGTGTCTTG 423

OY 147 GlyGluAlaProThrCysAspHisHisCysHisAsnHisLeuGlyLeuThrCysSer 166
|||||
Db 424 GGAGACTCAAGTCCCTGTGTGACCATTAATGTCCACACTACTTGGCGGCTACTCTCC 483

OY 167 CysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSerGly 186
|||||
Db 484 TGACAGAGCGGGCTACGTTCTCCACAGAACACACAGCGTCCACCTTTGTCAGGC 543

OY 187 GlnValPheThrGlnArgSerGlyGluLeuSerSerProGlyTyrProArgProTyrPro 206
|||||
Db 544 CAGGTGTTCACAGAGATCTGGGTATCTCAGTACGCTGAGTACCCACAGCATACCCA 603

OY 207 LysLeuSerSerCysThrTrpSerIleSerLeuGlnGlnGlyPheSerValIle-LeuAs 226
|||||
Db 604 AG-CTTCCAGCTGATCTTAAGATGCGCTGGAGAGGCTTCAGGGTCACTCTGGA 662

OY 226 PheValGlu-SerPheAspValGluThrHisProGlu-ThrLeuCysProTyrAspPhe 245
|||||
Db 663 CTTCGGGAGATCTTCGATGTGGGACGAGACCTGAAGGCCAAGTGGCCCTATGATGCG 722

OY 246 LeuLysIle----GlnThrAspArgGlnGlnHisGlyProPheCysGlyLysThrLeuPro 264
|||||
Db 723 CTCAGATTTTAAACAGACAGG-----GACACCGGCCCT---TGTGGAAAGCGCGGCC 773

BASE COUNT 121 a 111 c 97 g 146 t
 ORIGIN went through one round of normalization. Library constructed by Bento Soares and M. Patricia Bonaldo."

Alignment Scores:
 Pred. No.: 5,38e-77 Length: 475
 Score: 827.00 Matches: 155
 Percent Similarity: 98.10% Conservative: 0
 Best Local Similarity: 98.10% Mismatches: 3
 Query Match: 22.01% Indels: 0
 DB: 9 Gaps: 0

US-09-874-198-2 (1-686) x AA706982 (1-475)

QY 386 ThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPheThrMetLysVal 405
 475 ACCACCTACAAAGCTGTGATTCAGTACGTGGAAGAGACCTTTCACAAATGAAGTG 416
 406 AsnAspGlyLysTyrValCysGluAlaAspGlyPheTrpThrSerSerLysGlyLys 425
 415 AATGATGTAATATGTGTGTGAGCTGATGATTCGACGAGATCCAAAGAGAAAA 356
 QY 426 SerLeuProValCysGluProValCysGlyLeuSerAlaArgThrThrGlyArgIle 445
 355 TCACCTCCAGCTGTGAGCTGTTGTGTGACATACCCGACACAGAGAGCGGTATA 296
 QY 446 TyrGlyGlyGlnLysAlaLysProGlyAspPheProTrpGlnValLeuIleLeuGly 465
 295 TATGGAGGCAAAAGCAAAAGCTGTGATTTCTTGGCAAGCTCTGATTTAGGTGA 236
 QY 466 ThrThrAlaAlaGlyAlaLeuLeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaVal 485
 235 ACCACGACGACAGGTCTACTTTATATGACAACTGGGTCTTACACTGCTCATGCCGTC 176
 QY 486 TyrGlnGlnLysHisAspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeu 505
 175 TATGACCAAAAGCAAAAGCTGTGATTCCTGCAAGCTCTGATTTAGGTGA 116
 DB 175 TATGACCAAAAGCAAAAGCTGTGATTCCTGCAAGCTCTGATTTAGGTGA 116
 QY 506 SerProHisTyrThrGlnAlaTrpSerGluAlaValPheLeuHisGluGlyTrpHis 525
 115 TCACCTCATATACACACCTGTGTGAAAGCTGTTTATACATGAAGGTATATCTCAT 56
 QY 526 AspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnLysValVal 543
 55 GATGCTGGCTTTCACATGACATGACATGATTAATTTGAAACAAAGTTGTA 2
 DB 55 GATGCTGGCTTTCACATGACATGACATGATTAATTTGAAACAAAGTTGTA 2
 DEFINITION BF233497 752 bp mRNA linear EST 14-NOV-2000
 602024289F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4159415 5',
 mRNA sequence.
 ACCESSION BF233497
 VERSION BF233497.1 GI:11143969
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 752)
 NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM9437 row: b column: 24

FEATURES High quality sequence stop: 739.
 Location/Qualifiers
 1..752
 source

/organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4159415"
 /clone_lib="NCI_CGAP_L19"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 180 a 218 c 186 g 168 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.17e-74 Length: 752
 Score: 808.00 Matches: 145
 Percent Similarity: 88.40% Conservative: 15
 Best Local Similarity: 80.11% Mismatches: 21
 Query Match: 21.50% Indels: 0
 DB: 12 Gaps: 0

US-09-874-198-2 (1-686) x BF233497 (1-752)

QY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuCysGlySerValAlaThrProLeuGlyPro 20
 DB 41 ATGAGGCTACTCATCTTCCTGGGCTGCTGTGAGATTGGTGCCACACTTCTGGGTTCA 100
 QY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyTyrAla 40
 DB 101 AAGTGGCTGACACTGATTTGGGCGCTGTGTCTCCCTGTGCTCCACAGAAAGTGTCT 160
 QY 41 AsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
 DB 161 GACCATCAAGATGATGATCTGACAGCTGACAGCTGACCCCTGTGACCGCTGCGCTTAC 220
 QY 61 PheThrHisPheAspLeuGluLeuSerHisLeuCysGluTyrAspPheValLysLeuSer 80
 DB 221 TTCACCACTTGTACCTGGAACCTCTTACCCGCTGAGAGTATGACTTTGCAAGTTGAGC 280
 QY 81 SerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGlySerThrAspThrGluArgAla 100
 DB 281 TCAGGACCAAGGTCTGTGCGCACACTGTGTGCGCAGACAGTACACACTGACGCGCA 340
 QY 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
 DB 341 CTGGCAATGACACCTTCTACTGCTGCGGTCACAGCTTAAGGTACACTCCACTCGAC 400
 QY 121 TyrSerAsnGlnLysTrpThrGlyPheGlnAlaPheTyrAlaAlaGluAspIleAsp 140
 DB 401 TACTCCAAATGAGAAGCGCTTCACAGGCTTGAAGCCTTCTATGACACGGGAGTGTGAT 460
 QY 141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisCysHisAsnHisLeu 160
 DB 461 GAATGCAAGATGCTGTGGAAGCTAGTCCTGTGACCATTAATTGCCACAACACTCTTG 520
 QY 161 GlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgLysAspArgTrpCys 180
 DB 521 GGGGCTACTATGCTCGCGACAGGGGCTACGTTCTCCACAGAAACAAGCAGCTGCG 580
 QY 181 Ser 181
 DB 581 TCA 583
 RESULT 14
 AL022666/c 548 bp mRNA linear EST 29-DEC-1999
 LOCUS AL022666/1
 DEFINITION s7827b47 Beddington mouse dissected endoderm Mus musculus cDNA
 clone 528-1H14 5', mRNA sequence.
 ACCESSION AL022666
 VERSION AL022666.1 GI:6645253

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 548)
AUTHORS	Harrison,S.M., Dunwoodie,S.L., Arkell,R.M., Lehrach,H. and Beddington,R.S.
TITLE	Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo
JOURNAL	Development 121 (8), 2479-2489 (1995)
MEDLINE	95401865
COMMENT	Contact: Wiles,M., Lehrach,H. and Ayner,P. EBC Mouse Transcript Mapping Consortium Genoscope - CNS 2, rue Gaston Cremieux, 91000 Evry, France Email: pavner@pasteur.fr clone available from Ressourcenzentrum, Heubnerweg 6, D-14059 Berlin, Germany. Web site http://www.rzpd.de Seq primer: CCGTCCGGAATTCCTCCGGGT. High quality sequence only submitted. Vector: pSPori1; site_1: NotI; site_2: SalI; Cloned unidirectionally. Dissected endoderm 7.5 days. Average insert size: 1.2 kb (range: 0.2 - 2.kb). Location/Qualifiers 1..548 /organism="Mus musculus" /strain="C57BL/6 x DBA" /db_xref="taxon:10090" /clone="528-1H14" /clone_lib="Beddington mouse dissected endoderm" /tissue_type="dissected endoderm" /dev_stage="7.5 dpc" /note="Vector: pSPori1; site_1: NotI; site_2: SalI; Cloned unidirectionally. ~ High quality sequence only submitted. ~ Average insert size: 1.2 kb (range: 0.2 - 2.kb)"

BASE COUNT	133 a	142 c	117 g	156 t
ORIGIN				

Alignment Scores:			
Pred. No.:	1.67e-72	Length:	548
Score:	785.50	Matches:	144
Percent Similarity:	87.088	Conservative:	11
Best Local Similarity:	80.90%	Mismatches:	22
Query Match:	20.90%	Indels:	1
DB:	9	Gaps:	1

-874-198-2 (1-686) x AL022666 (1-548)	
502 Leu1y5atg1e5erProh15ytr7hrg1na1atp5erG1ua1aValp1he1h15glu	521
545 CTCAAAGGCGTCTCACTCATATACCTCAAGCGTGCCGAGAGAAATCTTTATACATGA	486
522 G1y1ytr7h15asp1aG1yph5asp5na5p11e1a1e1u11e1y5leu5na5n1y5	541
485 GCGTCACTCAAGCGGTGGTTTACATGATATATACATTATTAATCAAGACAAA	426
542 Val1a111eas1Ser5a5n11e7h1Pro11eG51e5a1uP1oArG1y5G1na1aG1u5erPh	561
425 GTCACATCTAACAGAGGCTTCATGCCGTGTTTCCCTACCGCCGAAAGAAAGACTGCATCTTA	366
562 MetArG1hr5p5p11e11y1hr1a5e5rG1y1rP1G1y1Leu1ThG1na5rG1y1p1he1u	581
365 ATGAAAGAGATTCACGTGACGACTGTGCGTGGGCTGTTAACCCAGAGGCGTCTT	306
562 AlaArGAsn1eum5e1y1Val5p11eP1o11eVal5pH15G1n1y5Cys1Th1a1a1a	601
305 GCTAAGAAACCTATGTGTGTGGACATATACCAATGTCTGCACCAAAATGTACACCGTG	246
602 Ty1rG1u1y5P1oP1oT1y1P1oArG1y5e1rVal1Th1a1a5n1e1u5Cys1a1aG1y1e	621

Db	245	TATGAAGAAG---CTCTATACGAGGAGTAAGAGTAAGACCGCTACACATGCTCTCTGCTGGCTTA	189
Qy	622	GLUSerGlyGlyLySASPserCysArgGlyLysSerpserGlyGlyAlaLeuAlaPheLeuAsp	641
Db	188	GAGACTGGTGGCAAGGAGACAGCTGCAGAGGTGACAGGTGGGGGGGATTAAGTCTTCTAGAT	129
Qy	642	SerGluThrGluATGTPPhPheValGlyGlyLeuAlaSerTPGlySerMetAsnCysGly	661
Db	128	AATGACACACAGCCATGTTTGTGTGGAGAGATTAAGTTTCTTCTGGGTTCCATTAATTTGGCG	69
Qy	662	GLUAlaGlyGlnTyrglyValTyThrLysValIleAsnTyrlIleProTyrIle	679
Db	68	GGCGACAGCAGTATGGGGCTACACAAAGTCACTACACTATATTCCTCGATT	15
RESULT 15			
LOCUS	A1529228		
DEFINITION	A1529228	586 bp	RNA linear EST 18-MAR-1999
ACCESSION	A1529228		
VERSION	A1529228		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Enkaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 586) Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.		
TITLE	The Washu-NCI Mouse EST Project 1999		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Marra M/washu-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@washington.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:9711221		
FEATURES	Possible reversed clone: similarity on wrong strand Seq primer: custom primer used High quality sequence stop: 501.		
SOURCE	Location/Qualifiers 1. 586 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1886897" /clone_1lb="Sugano mouse liver mlia" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: liver; Vector: pME18S-FLG3; Site:1: DraIII (CACTGATG); Site:2: DraIII (TGTGGGCTCACTG); 1st strand cDNA was primed with an oligo(dt) primer (ATGTGGCTTTTCTTTTCTTTT); double-stranded cDNA was ligated to a DraIII adaptor (TGTGGGCTCACTG); digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGATG, 3' site CACCATG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTTAAGCTGG and 3' end primer CGACTCGACTCGACACA."		
BASE COUNT	125 a	167 c	155 g
ORIGIN			137 t
			2 others

DB 843 GAGAAAGATTGGCACTTCCTCCGTAACATCCGATGCG---CATCTCCAAAGATTCTCACC 787
OY 508 HlATyThrGlnAlaTrpSerGluAlaValPheLleHisGluGlyTrpThrHisAspAla 527
DB 786 CAT---ACATCAAGCTGGCCCGAGGA-ATCTTAGACATGAAGCTTACACTCAGCGTGCT 731
OY 528 GtYPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValIleAsnSerAsn 547
DB 730 GGTATTGACATGATATAGATTGATTAACTCAAGACAAACTCACAATCAACGAGAGC 671
OY 548 IleThrProIleCysLeuProAlaGlyGluAlaGluSerPheMetArgTrpAspAspIle 567
DB 670 ATCATGCTGTTGGCTTACCGCGCAAAAGAGCTCATCTTATGAGAACAGACTTCAC 611
OY 568 GtYThAlaSerGlyTrpGlyLeuThrGlnAlaGlyPheLeuAlaTrpAsnLeuMetYr 587
DB 610 GGAACTGTGGCTGGCTGGGGGTTACCCAGAGGGGCTTCTCTGAAACCTAAATGTTT 551
588 ValAspIleProIleValAspHisGlnLysCysThrAlaAlaTrpGluLysProProTrp 607
550 GTGGACATACCAATTGCTGTGACCAACAAATGATACCGCGCTGATGAAG---CTCTAT 494
OY 608 ProAlaGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGlyGlyAsp 627
DB 493 CCAGGAGTAAGAGTACGCGCTTACATGCTCTGCTGCTTACAGATGCTGGCAAGGAC 434
OY 628 SerCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluProGluArgTrp 647
DB 433 AGCTGCAGAGGTACAGTGGGGGCGATTGATTTCTTCAATATGAGACACAGCGATGG 374
OY 648 PheValGlyGlyIleValSerTrpGlySerMetAsnCysGlyGluAlaGlyGlnTrpGly 667
DB 373 TTGTGGGAGGAGTATGTTCTCTGGGGTTCATTAATTGTGGGGCGGAGACACGATAGG 314
OY 668 ValTrpThrLysValIleAsnTrpIleProTrpIleGluAsnIleIleSerAspPhe 666
DB 313 GTCTACACAAAGTCACTATATTTCCCTGGATTGAGACATATATTAATTTTC 257
RESULT 19
AA238260
LOCUS 552 bp mRNA linear EST 03-MAR-1997
DEFINITION mx15a01.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:680232 5'
RA-REACTIVE FACTOR PRECURSOR ; mRNA sequence.
ACCESSION
AA238260
AA238260.1 GI:1862301
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 552)
Mammals: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
Marta,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MGI:419936
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 423.

FEATURES
source Location/Qualifiers
1..552
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:680232"
/clone_lib="Soares mouse NML"
/issue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7p3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTATCCATCTGGAAGTGGAGCGGCCGCGAATCTTTTATTTTATTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7p3 vector. Library
constructed and normalized by Bento Soares and M.Falima
Bonaldo."
BASE COUNT 111 a 161 c 149 g 130 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1 37e-69 Length: 552
Score: 758.00 Matches: 138
Percent Similarity: 86.78% Conservative: 13
Best Local Similarity: 79.31% Mismatches: 23
Query Match: 20.17% Indels: 0
DB: 9 Gaps: 0
US-09-874-198-2 (1-686) x AA238260 (1-552)
OY 1 MetArgLeuLeuThrLeuLeuGlyLeuLysGlySerValAlaThrProLeuGlyPro 20
DB 31 ATGAGGCTACTCATTCCTTCCTGGGCTGCTGTGGAGTTTGTCGACACACTTCGGGTCA 90
OY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyTrpAla 40
DB 91 AAGTGGCTGTAACCTGATATTCGGGCGCTGTCTCCCTGGCTTCCACAGAGATGTGCT 150
OY 41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTrpArgLeuArgLeuTrp 60
DB 151 GACCATCAAGATGATCTGTGACACTGACATGACACCCCTGGTACCGCCCTGCTTAC 210
OY 61 PheThrHisPheAspLeuGluLeuSerHisLeuCysGlyTrpArgPheValLysLeuSer 80
DB 211 TTACACCACCTTGACCTGGAGACCTCTTACCGCTGCGAGATGACTTTTGCAAGTTGAGC 270
OY 81 SerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAla 100
DB 271 TCAGGCGCCAGGCTGCTGGCCACACTGTGTGGCCAGAGAGTACAGACACTGAGAGCA 330
OY 101 ProGlyLysAspThrPheTrpSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
DB 331 CCTGGCAATGACACTTCTACTCTGAGTGGTCCAGGCTTAAAGGTCACTTCCACTCCGAC 390
OY 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTrpAlaAlaGluAspIleAsp 140
DB 391 TACTCCAAATGAGAGCGCTTACAGAGGTTTGAGAGCCCTTATGACAGCGAGATGCGAT 450
OY 141 GluGlyGlnValAlaProGlyGluAlaProThrCysAspHisCysHisAsnHisLeu 160
DB 451 GAATGCAAGATGCTCTGTGGAGACACGCTCTGTGACCATTATTTGCCACAACCTACTTG 510
OY 161 GtYGLyPheTrpCysSerCysArgAlaGlyTrpValLeuHis 174
DB 511 GGGCGGCACTANTGCTCTGCAAGAGGGGCTACGTTCTCCAC 552
RESULT 20
BO644570 939 bp mRNA linear EST 15-JUL-2002
LOCUS BO644570
DEFINITION AGENCOUNT_8490607 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6293770
5', mRNA sequence.
ACCESSION BO644570
VERSION BO644570.1 GI:21768742

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 939)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM2497 row: f column: 11
High quality sequence stop: 551.
Location/Qualifiers
RES 1..939
source /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6293770"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."
BASE COUNT 189 a 308 c 273 g 166 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 2.06e-67 Length: 939
Score: 741.00 Matches: 161
Percent Similarity: 67.76% Conservative: 5
Best Local Similarity: 65.71% Mismatches: 25
Query Match: 19.72% Indels: 54
DB: 14 Gaps: 9
9-874-198-2 (1-686) x BQ644570 (1-939)
1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
20 ATAGAGGCTGTGACCTCTCGGGCTTCTGTGTGGCTGGTGGCCACCCCTTGGGCGG 79
21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAla 40
80 AAGTGGCTGAACCTGTGTGGGGCTGGATGCCCGCGCTTCCAGGGAGTATGCC 139
41 AsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
140 AATGACCAAGAGGGGGCTGGACCTGACAGCACCCCGCGACACCGCCCTCTTAC 199
61 PheThrHisPheAspLeuGluLeuSerHisLeuGlyGluTyrAspPheValTyrLeuSer 80
200 TTACACCACTTCACCTGGAGCTCTCCACCTCTGGAGAGTACGACTTCGTAAGCTGAGC 259
81 SerGlyAlaValLeuAlaThrLeuGlyGlyGlnGluSerHisThrAspThrGluArgAla 100
260 TCGGGGGCCAAAGTGTGGGCGACGCTGTGGGGCAGAGAGACAGACAGAGGCGGCC 319
101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
320 CCGGCAAGACACTTCTACTGCTGGGGCTCCAGCCTGACATTAACCTTCGCTCGAC 379

121 TyrSerAsnGluLysProPheThrGlnPheGluAlaPheTyrAlaAlaLysPheAsp 140
380 TACTCCACAGACAGCCGTTCCAGGGGTTGAGGCTTGTATGACCCAGCCAGGTAGCCA 439
141 Glu-----CysGlnVal 144
440 AGAGGGGTCCTGCACATGTCAGTGTGGCAGGTGGCTGTGGGGTAACTGTGT---CTT 496
145 AlaPro-----GlyAlaAlaProGln 151
497 AGGCCAGGAGAGCCCTGCTCACTTCCACCTTTCCAGGGCAGGAGGAGGCTCTG 556
152 CysAspHisHis-----CysHisAsnHisLeuGly 161
557 GCGTGAATCATGTCACAAAGACAAAGACAGCCGTCAGCTCATTCACAT----- 610
162 GlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgTrpCysSer 181
611 GGGCGTAGTGGCCAACTGTGAGC-----AGGATGTGAGACAGCATCG 655
182 AlaLeuGlySerGlyValAlaPheThrGlnArgSerGlyGluLeuSer-----SerPro 199
656 CCG-----CAGAGT---ACGAGGGGAGTGGCGGGCGCAGCAGCTCAGCCCT 700
200 GluTyrProArgPro 204
701 GTAATTCAGACACT 715
RESULT 21
BQ653181 1035 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8303990 NIH_MGC_100 Homo sapiens cDNA IMAGE:6269223
DEFINITION 5', mRNA sequence.
ACCESSION BQ653181
VERSION BQ653181.1 GI:21777353
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1035)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM2442 row: g column: 16
High quality sequence stop: 510.
Location/Qualifiers
FEATURES
source 1..1035
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6269223"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

BASE COUNT 224 a 341 c 298 g 172 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5 01e-67 Length: 1035
 Score: 738.00 Matches: 157
 Percent Similarity: 70.26% Conservative: 6
 Best Local Similarity: 67.67% Mismatches: 28
 Query Match: 19.64% Indels: 42
 DB: 14 Gaps: 4

US-09-874-198-2 (1-686) x BQ653181 (1-1035)

Oy 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
 |||||
 Db 20 ATGAGGCTGCTGACCTCTGGGCGCTTGTGTGGCTCGGTGGCCACCCCTGGGCGCG 79

21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyLysTyrAla 40
 |||||
 Db 80 AAGTGGCCGTAACCTGTGTCGGGCGCTGGGCAATCCCGGCTTTCACGGGAGTATGCC 139

Oy 41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
 |||||
 Db 140 AATGACACGAGAGCGGCGCTGGACCTGACTGCACCCCGGCTACCGCGCTGCTCTAC 199

Oy 61 PheThrHisPheAspLeuGluLeuSerHisLeuGlyLysTyrAspPheValLysLeuSer 80
 |||||
 Db 200 TTCACCCACTGTGACCTGGAGGCTCCACACTGGGAGTACGACTTCGTCACAGCTGAGC 259

Oy 81 SerGlyAlaLysValLeuAlaThrLeuGlyGlnGlnSerThrAspThrGluTyrAla 100
 |||||
 Db 260 TCGGGGGCCAAAGGTGTGGCCAGGCTGTGGGCGAGAGGCAACACAGGAGCGGGCC 319

Oy 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspThrPheArgSerAsp 120
 |||||
 Db 320 CCTGGCAAGACACTTCTACTGCTGGGCTCCAGGCTTGACATTACCTTCCCTCCGAC 379

Oy 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaAlaGluAsp-11leas 140
 |||||
 Db 380 TACTCCAAACGAGAGCGGTTTCAGGCGGTTTCAGAGGCTTCTATGACCGAGGTTGAGCCA 439

Oy 140 PGLuCySGlnValAlaProGlyGluAlaProThrCysAspHisHisAsnHisLe 160
 |||||
 Db 440 AGAG----- 443

Oy 160 uGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrcy 180
 |||||
 Db 444 -----GGTCCCTGC-----AACATCTCACTCTG 466

180 sSerAlaLeuGlySerGly---GlnValPheThrGlnArgSerGlyGluLeuSerSerP 199
 |||||
 Db 467 CGGAGCTGGCTGTGGGGTAACCTGTCTTAGGCCAGGC-AGCCCGGCTTCACGTTTCCC 525

Oy 199 cGluTyrProArgPro-----TyrProLysLeuSerSerCy 211
 |||||
 Db 526 CACCTTTCCCGAGGCGAGGAGAGGCGCTGCGCTGACATCATCATGCAAGACCA 585

Oy 211 sThrTyrSerIleSerLeuGluGluGlyPheSer 222
 |||||
 Db 586 AACAGCGCGGAGCTCCATTTCATGGGCTGAGCT 619

RESULT 22
 BQ648641 913 bp mRNA linear EST 15-JUL-2002
 LOCUS BQ648641 AGENCOURT 8299192 NIH_MCC_100 Homo sapiens CDNA clone IMAGE:6271535
 DEFINITION 5' mRNA sequence.
 ACCESSION BQ648641
 VERSION BQ648641.1 GI:21772813
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 913)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2448 row: h column: 04
 High quality sequence stop: 549.
 Location/Qualifiers

FEATURES
 source
 1. 913
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6271539"
 /clone_lib="NIH_MGC_100"
 /issue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 193 a 298 c 255 g 166 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 6.63e-67 Length: 913
 Score: 736.00 Matches: 155
 Percent Similarity: 73.15% Conservative: 3
 Best Local Similarity: 71.76% Mismatches: 25
 Query Match: 19.58% Indels: 34
 DB: 14 Gaps: 3

US-09-874-198-2 (1-686) x BQ648641 (1-913)

Oy 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
 |||||
 Db 20 ATGAGGCTGCTGACCTCTGGGCGCTTGTGTGGCTCGGTGGCCACCCCTGGGCGCG 79

Oy 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyLysTyrAla 40
 |||||
 Db 80 AAGTGGCCGTAACCTGTGTCGGGCGCTGGGCAATCCCGGCTTTCACGGGAGTATGCC 139

Oy 41 AsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
 |||||
 Db 140 AATGACACGAGAGCGGCGCTGGACCTGACTGCACCCCGGCTACCGCGCTGCTCTAC 199

Oy 61 PheThrHisPheAspLeuGluLeuSerHisLeuGlyLysTyrAspPheValLysLeuSer 80
 |||||
 Db 200 TTCACCCACTGTGACCTGGAGGCTCCACACTGGGAGTACGACTTCGTCACAGCTGAGC 259

Oy 81 SerGlyAlaLysValLeuAlaThrLeuGlyGlnGlnSerThrAspThrGluTyrAla 100
 |||||
 Db 260 TCGGGGGCCAAAGGTGTGGCCAGGCTGTGGGCGAGAGGCAACACAGGAGCGGGCC 319

Oy 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspThrPheArgSerAsp 120
 |||||
 Db 320 CCTGGCAAGACACTTCTACTGCTGGGCTCCAGGCTTGACATTACCTTCCCTCCGAC 379

Oy 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaAlaGluAsp-11leas 140
 |||||
 Db 380 TACTCCAAACGAGAGCGGTTTCAGGCGGTTTCAGAGGCTTCTATGACCGAGGTTGAGCCA 439

OY	140	pclucylglnvalalaprogllylualalaprothrhcysasprshstiscysshiasnhi5le	160
Db	440	ACAG-----	443
OY	160	wclglypheryCySSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCy	180
Db	444	-----GGGTCTCC-----AACATCTCAGTCTCG	466
OY	180	sSerAlaLeuCySserGly---GlnValPheThrIlnarGserGlyGluLeuSerSerPr	199
Db	467	CccAcctGgcTcTgtGggGgGtAAcCTcTgTCTTAGCCAGGC-ACCcCTGcCTTCAGTTCC	525
OY	199	oGlUurYrProArGyProTYrProLYsLeuSerSerCySThyrSer	214
Db	526	CACCTTCCCAggGcAgGgNGAGAGcCTCTGcCTGACATCATCC	571
RESULT 23			
BQ648719			
LOCUS			
DEFINITION	BQ648719	921 bp	mRNA linear EST 15-JUL-2002
LOCATION	AGNCOURT 8188306 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283375		
KEYWORDS	5' mRNA sequence.		
SOURCE	BQ648719		
ORGANISM	BQ648719.1 GI:21772891		
	EST.		
	human.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 921)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CM2479 row: e column: 08 High quality sequence stop: 609. Location/Qualifiers 1..921 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image:6283375" /clone_lib="NIH_MGC_100" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: liver; Vector: pOTf7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."		
FEATURES			
source			
BASE COUNT	172 a	300 c	285 g
ORIGIN			164 t
Alignment Scores:			
Pred. No.:	6.73e-67	Length:	921
Score:	736.00	Matches:	149
Percent Similarity:	73.91%	Conservative:	4
Best Local Similarity:	71.98%	Mismatches:	19
Query Match:	19.58%	Indels:	35
DB:	14	Gaps:	3

QY	1	MetaGlaLeu	enThLeu	enClyLeu	enCysGlySerValAla	thProLeu	ClyPro	20
Db	20	ATGAGGCTGCTGACCTCTGGGCTTGTGTGTGCTCGTGCCACCCCTTGGGCGCG	79					
QY	21	LysTTPProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyLysVal	40					
Db	80	AAGTGGCTGAACCTGTGTCGGGGCTGGCATCCCGCGCTTCCAGGGAGATATGCC	139					
QY	41	AsnAspGlnGluArgArgTTPThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr	60					
Db	140	AATGACCCAGGAGGCGCTGGACCTGTGACTGACCCCGCGGTACCGCTCGGCTCTTAC	199					
QY	61	PheThrHisPheAspLeuGlyIleuSerHisLeuCysGlyTyrAspPheValLysLeuSer	80					
Db	200	TTTCACCCACTTGACCTGGAGTGCTGCCACCTGTCCAGTAGTACACTTGTCAAGCTGAGC	259					
QY	81	SerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAla	100					
Db	260	TGCGGGGCCAAGTGCTCTGGCACGCTGTGGGGCACGAGACACAGACGAGGAGGCGCC	319					
QY	101	ProGlyLysAspThrPheTyrSerLeuGlySerLeuAspIleThrPheArgSerAsp	120					
Db	320	CCTGGCAAGACACTTCTTACTCGCTGGGCTTCAGACCTGGACATTACCTTCGCTCGAC	379					
QY	121	TyrSerAsnGluLysProPheThrGlyPheGlnAlaPheTyrAlaAlaGluAsp-IleAs	140					
Db	380	TACTCCACGAGAACCCGTCACGGGGTTGAGGCGCTCTATGACGCGAGGAGTGAGCCA	439					
QY	140	PcLysGlnValAlaProGlyGluAlaProThrCysAspHisIleCysHisAsnHisLe	160					
Db	440	AGAG-----	443					
QY	160	uGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCy	180					
Db	444	-----GGGTCTGTC-----	466					
QY	180	sSerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGly	200					
Db	467	CGCAGCTGGCTGGGGGTACTCTGTCTTA-----GGCCAGGACGCCCTGCCTTC	517					
QY	200	uTyrProArgProTyrPro	206					
Db	518	AGTTCCCCACCTTCCCA	536					
RESULT	24							
LOCUS	B0646367	975 bp	mRNA	linear	EST 15-JUL-2002			
DEFINITION	AGNCOURT 8488545 NIH_MGC_100 Homo sapiens cdna clone IMAGE:6286149							
ACCESSION	B0646367	5' mRNA sequence.						
VERSION	B0646367.1	GI:21770539						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eubacteria; Primates; Catarrhini; Hominiidae; Homo.							
REFERENCE	1 (bases 1 to 975)							
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .							
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)							
JOURNAL	Unpublished (1999)							
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapds-remail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L10CM2503 row: i column: 14 High quality sequence stop: 504.							
FEATURES	1..975							
source	location/Qualifiers							

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6296149"
/clone.lib="NIH_MGC_100"
/tissue.type="hepatocellular carcinoma, cell line"
/lab.host="DH10B (phage-resistant)"
/Note="Organ: Liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
BASE COUNT      198 a      320 c      286 g      171 t
ORIGIN
...
ment Scores:
  e:          7.39e-67      Length:      975
Percent Similarity: 73.6.00      Matches:      149
Best Local Similarity: 73.91%      Conservative: 4
Query Match:      19.98%      Mismatches: 19
DB:              14          Indels:      35
Gaps:            3
US-09-874-198-2 (1-686) x B0646367 (1-975)
OY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
Db 20 ATGAGGCTGTGACCTCTGGGCTGTGTGTGGCTGGGCAACCCCTTGGGCGG 79
OY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAla 40
Db 80 AAGTGGCCGAACCTGTGTGGGCGGCTGGCATCCCGGCTTCCAGGGAGTATGCC 139
OY 41 AsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
Db 140 AATGACACAGAGGCGGCTGGACCTGACTGCACCCCGGCTACCGGCTGCTTAC 199
OY 61 PheThrHisPheAspLeuLeuLeuSerHisLeuGlyGluTyrAspPheValLysLeuSer 80
Db 200 TTCACCCACTTGACCTGGAGGCTCTCCACCTCTGCGAGTACGACTTCGTAAGCTGAGC 259
OY 81 SerGlyAlaLysValLeuAlaThrLeuLeuGlyGlnGlySerThrAspThrGluArgAla 100
Db 260 TCGGGGGCCAAAGGTGCTGGGCAAGCTGTGGGCGAGAGACACACAGGCGGGCC 319
OY 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspThrPheArgSerAsp 120
Db 320 CCTGGCAAGACACTTCTTACTCGCTGGGCTCCAGCCTGACATTCCTCCGCTCCGAC 379
OY 121 TyrSerAsnGlnLysProPheThrGlyPheGluAlaPheTyrTrpAlaGluAsp-IleAs 140
Db 380 TACTCCAAAGAGAGCCGTTTCAGGGGTTGAGAGCCCTTATGCAACCCGAGGTAGCCA 439
OY 140 PclLysGlnValAlaProGlyLysValaLarProThrCysAspHisHisnHisIle 160
Db 440 AAGAG-----
OY 160 uclGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgTrpCys 180
Db 444 -----GGGTCTCTGC-----AACATCTCACTGTC 466
OY 180 sSerAlaLeuGlySerGlyGlnValaPheThrGlnArgSerGlyGlnLeuSerSerProG 200
Db 467 CCGAGCTGGCTGTGGGGGTAACCTGTCTTTA-----GGCCAGGCAAGCCCTGCTTC 517
OY 200 uTyrProArgProTyrPro 206
Db 518 AGTTTCCCAACCTTTCCTCA 536
RESULT 25

```

```

B0650766
LOCUS      B0650766      796 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT.8347134 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284012
5', mRNA sequence.
ACCESSION B0650766
VERSION   B0650766.1 GI:21774938
KEYWORDS
SOURCE
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 796)
  NIH-MGC http://mhc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: CGAP (Stanford)
  cDNA Library Preparation: Rubin Laboratory
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  DNA distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: L1CM2480 row: 0 column: 21
  High quality sequence stop: 505.
  Location/Qualifiers
    1..796
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:6284012"
      /clone.lib="NIH_MGC_100"
      /tissue.type="hepatocellular carcinoma, cell line"
      /lab.host="DH10B (phage-resistant)"
      /Note="Organ: Liver; Vector: pOTB7; Site_1: XhoI; Site_2:
      EcoRI; cDNA made by oligo-dT priming. Directionally cloned
      into EcoRI/XhoI sites using the following 5' adaptor:
      GGCACGAG(G). Size-selected >500bp for average insert size
      1.8kb. Library constructed by Ling Hong in the laboratory
      of Gerald M. Rubin (University of California, Berkeley)
      using ZAP-cDNA synthesis kit (Stratagene) and Superscript
      II RT (Life Technologies). Note: this is a NIH_MGC
      library."
BASE COUNT      142 a      264 c      243 g      144 t      3 others
ORIGIN
...
Alignment Scores:
  Pred. No.:      6.76e-67      Length:      796
Score:          735.00      Matches:      137
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      19.56%      Indels:      0
Gaps:            0
US-09-874-198-2 (1-686) x B0650766 (1-796)
OY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuGlySerValAlaThrProLeuGlyPro 20
Db 32 ATGAGGCTGTGACCTCTGGGCTGTGTGTGGCTGGGCAACCCCTTGGGCGG 91
OY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAla 40
Db 92 AAGTGGCCGAACCTGTGTGGGCGGCTGGCATCCCGGCTTCCAGGGAGTATGCC 151
OY 41 AsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
Db 152 AATGACACAGAGGCGGCTGGACCTGACTGCACCCCGGCTACCGGCTGCTTAC 211
OY 61 PheThrHisPheAspLeuLeuSerHisLeuGlyGluTyrAspPheValLysLeuSer 80
Db 212 TTCACCCACTTGACCTGGAGGCTCTCCACCTCTGCGAGTACGACTTCGTAAGCTGAGC 271
OY 81 SerGlyAlaLysValLeuAlaThrLeuLeuGlyGlnGlySerThrAspThrGluArgAla 100

```

```

Db      272  TCGGGGGCCAGGTGCTGCGCAGCGGCGAGAGACAGACAGAGCGGCGGCC 331
QY      101  ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
Db      332  CCGGCAAGGACACTTCTACTGCTGGGCTCCAGCGCTGACATTACCTCCGCTCCGAC 391
QY      121  TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaAlaGlu 137
Db      392  TACTCCACGAGAGCGCTTCACGGGGTTCGAGGCGCTTCTATGACGCGGAG 442

```

Search completed: January 24, 2003, 00:17:05
 Job time : 2236 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 23, 2003, 23:41:12 : Search time 126 Seconds

(without alignments)
2445.997 Million cell updates/sec

Title: US-09-874-198-2

Perfect score: 3758

Sequence: 1 MRLTLGLGCGSVATPLGP.....GYTVKVINYPWNIENISDF 686

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 75 summaries

Command line parameters:

-MODEL=frame_plus.model -DEV=xlp
-Q/cgcn2_1/USPTO_SPOOL/US09874196/runat.22012003.085003.10992/app_query.fasta.1.839
-DB=Published_Applications_MA -QMT=fastlap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62
-TRANS=human40.cdi -LIST=75 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALLIGN=25 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USPR=US09874198 -CGCN_1.1.36 -runat.22012003.085003.10992
-NCPU=6 -ICPU=3 -NO_XLPEXT -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgcn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgcn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgcn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgcn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgcn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgcn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgcn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgcn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgcn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgcn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgcn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgcn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgcn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgcn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3758	100.0	2475	10 US-09-874-198-3	Sequence 3, Appli
2	3758	100.0	2475	10 US-09-874-238-3	Sequence 3, Appli
3	1412	37.6	2493	10 US-09-880-107-2256	Sequence 2256, Ap
4	1412	37.6	2555	12 US-10-044-090-613	Sequence 613, App

5	1410	37.5	2386	9 US-09-808-602-92	Sequence 92, Appl
6	1238	32.9	2647	10 US-09-880-107-2105	Sequence 2105, Ap
7	1238	32.9	2659	12 US-10-044-090-472	Sequence 472, App
8	1238	32.9	2725	10 US-09-925-301-182	Sequence 182, App
9	1214	32.3	2908	10 US-09-917-800A-1348	Sequence 1348, Ap
10	873	23.2	2063	10 US-09-925-302-169	Sequence 169, App
11	689.5	18.3	1148	9 US-09-925-302-168	Sequence 168, App
12	498	13.3	3318	9 US-09-808-602-91	Sequence 91, Appl
13	497	13.2	1356	9 US-09-808-602-20	Sequence 20, Appl
14	496	13.2	1667	9 US-09-808-602-16	Sequence 16, Appl
15	496	13.2	1691	9 US-09-808-602-18	Sequence 18, Appl
16	490	13.0	3147	9 US-10-099-700A-1	Sequence 1, Appl1
17	490	13.0	3147	9 US-10-099-700A-3	Sequence 3, Appl1
18	481	12.8	2756	10 US-09-925-301-351	Sequence 351, App
19	469.5	12.5	3106	10 US-09-900-751-1	Sequence 1, Appl1
20	446.5	11.9	807	10 US-09-981-123-1	Sequence 1, Appl1
21	444	11.8	2409	10 US-09-888-615-54	Sequence 54, Appl1
22	444	11.8	3143	9 US-09-978-295A-168	Sequence 168, App
23	444	11.8	3143	9 US-09-978-697-168	Sequence 168, App
24	444	11.8	3143	9 US-09-978-192A-168	Sequence 168, App
25	444	11.8	3143	9 US-09-999-832A-168	Sequence 168, App
26	444	11.8	3143	9 US-09-978-189-168	Sequence 168, App
27	424.5	11.3	2913	10 US-09-888-615-42	Sequence 42, Appl1
28	416.5	11.1	1327	9 US-09-978-295A-170	Sequence 170, App
29	416.5	11.1	1327	9 US-09-978-697-170	Sequence 170, App
30	416.5	11.1	1327	9 US-09-978-192A-170	Sequence 170, App
31	416.5	11.1	1327	9 US-09-999-832A-170	Sequence 170, App
32	416.5	11.1	1327	9 US-09-978-189-170	Sequence 170, App
33	411	10.9	3919	10 US-09-942-366-4	Sequence 4, Appl1
34	410	10.9	2792	10 US-09-118-748-1	Sequence 1, Appl1
35	405	10.8	1413	10 US-09-884-901-2	Sequence 2, Appl1
36	405	10.8	1548	10 US-09-150-811-6	Sequence 6, Appl1
37	403.5	10.7	2306	9 US-10-004-551-3	Sequence 3, Appl1
38	403.5	10.7	2846	9 US-09-992-596-230	Sequence 230, App
39	403.5	10.7	2846	9 US-09-989-722-230	Sequence 230, App
40	403.5	10.7	2846	9 US-09-989-732-230	Sequence 230, App
41	403.5	10.7	2846	9 US-10-063-514-37	Sequence 37, Appl
42	403.5	10.7	2846	9 US-09-989-735-230	Sequence 230, App
43	403.5	10.7	2846	9 US-09-989-740-230	Sequence 230, App
44	403.5	10.7	2846	9 US-09-989-436-230	Sequence 230, App
45	403.5	10.7	2846	9 US-09-991-181-230	Sequence 230, App
46	403.5	10.7	2846	9 US-09-993-687-230	Sequence 230, App
47	403.5	10.7	2846	9 US-09-989-734-230	Sequence 230, App
48	403.5	10.7	2846	9 US-09-997-653-230	Sequence 230, App
49	403.5	10.7	2846	9 US-10-174-590-169	Sequence 169, App
50	403.5	10.7	2846	9 US-10-176-758-169	Sequence 169, App
51	403.5	10.7	2846	9 US-10-063-616-37	Sequence 37, Appl
52	403.5	10.7	2846	9 US-09-989-722-230	Sequence 230, App
53	403.5	10.7	2846	10 US-09-989-732-230	Sequence 230, App
54	403.5	10.7	2846	10 US-09-989-733-230	Sequence 230, App
55	403.5	10.7	2846	10 US-09-989-729-230	Sequence 230, App
56	403.5	10.7	2846	10 US-09-989-727-230	Sequence 230, App
57	403.5	10.7	2846	10 US-09-989-731-230	Sequence 230, App
58	403.5	10.7	2846	10 US-09-989-732-230	Sequence 230, App
59	403.5	10.7	2846	10 US-09-991-073-230	Sequence 230, App
60	403.5	10.7	2846	10 US-09-990-442-230	Sequence 230, App
61	403.5	10.7	2846	10 US-09-991-163-230	Sequence 230, App
62	403.5	10.7	2846	10 US-09-993-604-230	Sequence 230, App
63	403.5	10.7	2846	10 US-09-990-456-230	Sequence 230, App
64	403.5	10.7	2846	10 US-09-989-721-230	Sequence 230, App
65	403.5	10.7	2846	12 US-10-006-867-37	Sequence 37, Appl
66	403.5	10.7	2846	12 US-10-052-586-169	Sequence 169, App
67	403.5	10.7	3546	10 US-09-850-048A-3	Sequence 3, Appl1
68	398.5	10.6	1059	10 US-09-888-615-36	Sequence 36, Appl1
69	398	10.6	2553	10 US-09-888-615-49	Sequence 49, Appl1
70	397	10.6	1543	10 US-09-917-800A-1575	Sequence 1575, Ap
71	396.5	10.6	5021	10 US-09-285-385C-1	Sequence 1, Appl1
72	395.5	10.5	1783	10 US-09-919-048-188	Sequence 188, App
73	395	10.5	2363	10 US-09-880-107-3690	Sequence 3690, Ap
74	385	10.5	4661	10 US-09-285-385C-3	Sequence 3, Appl1
75	394	10.5	2457	10 US-09-850-048A-1	Sequence 1, Appl1

ALIGNMENTS

```
RESULT 1
US-09-874-198-3
; Sequence 3, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USFS FOR IT
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; ID NO 3
; NCBI: 2475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)...(2094)
US-09-874-198-3

Alignment Scores:
Pred. No.: 0 Length: 2475
Score: 3758.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-874-198-2 (1-686) x US-09-874-198-3 (1-2475)
QY 1 MetArgLeuThrLeuLeuGlyLeuLeuCysGlySerValAlaThrProLeuGlyPro 20
Db 37 ATGAGGCTGTCACCTCTCGGGCTTCTGTGGCTGGTGGCCACCCCTTAGGCGCG 96
QY 21 LysTrpProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGlyValAla 40
Db 97 AAGGGGCTGAACCTGTGTGGGGCGCTGCAATCCCGCGCTTCCAGGGAGTATGCC 156
QY 41 AsnAspGlnIuArgArgTrpThrLeuThrAlaProProGlyTyrArgLeuArgLeuTyr 60
Db 157 AATGACCAAGAGGCGGCTGGACCTGACTGACACCCCGGCTACCGCTGCGCTCTAC 216
QY 61 PheThrHisPheAspLeuGluLeuSerHisLeuCysGlyIuThrAspPheValLysLeuSer 80
Db 217 TTCACCCACTTCAGCTCGAGGCTCTCCACCTCTCGAGAGTACACATTCGCAAGCTGAGC 276
QY 81 SerGlyAlaValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAla 100
Db 277 TCGGGGCGCAAGGTGCGGCGCACGTGTGGCGAGAGACACAGACAGCGAGCGGGCC 336
QY 101 ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
Db 337 CCTGGCAAGGACACTTTCATCTGCTGGGCTCGACGCTGACATTTACCTTCCGCTCGAC 396
QY 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaIleAspIleAsp 140
Db 397 TACTCCACAGAGAAGCGTTTCAGGGGTTTGAGAGCCTTCTATGTCAGCGGAGCATTTGAC 456
QY 141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisIscysHisAsnHisLeu 160
Db 457 GAGTGCAGGTGGCGCGGAGAGGCGCCACCTGCGACACCAACCACTGCCAACCACTG 516
QY 161 GlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
Db 161 GlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
```

```
Db 517 GCGGCTTCTACTGCTCTGCGCGGAGGCTACGCTGACCCGTAAACAGCGCACCTGC 576
QY 181 SerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyLeuLeuSerProGlu 200
Db 577 TCAGCCCTGTGCTCCGGCCAGGCTCTCACCCAGAGGTGTGGAGCTGACGACCCCTGAA 636
QY 201 TyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGly 220
Db 637 TACCACGGGCGGTATCCCAAACTCTCAGTTGACATTCAGCATCAGCTCGAGGAGGG 696
QY 221 PheSerValIleLeuAspPheValIuSerPheAspValGluThrHisProGluThrLeu 240
Db 697 TTGAGTGTATTCGTGACTTGTGGAGTCTTGAGTGTGAGACACACCTGTAACCTG 756
QY 241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGly 260
Db 757 TGTCCCTACGACTTCTCAAGATTCAACAGACAGAAAGAACATGGCCATTCTGTGG 816
QY 261 LysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIleThrPheValThr 280
Db 817 AAGACATTCGCCCCACAGATTGAACAAAGAACACAGCGGTGACCATCTGTTCACA 876
QY 281 AspGluSerGlyAspHisThrGlyTyrLysIleHisTyrThrSerThrAlaGlnProCys 300
Db 877 GATGAATCAGAGACACACAGCTGGAAGATCCACTACAGCAGCAGCAGCGACCTTGC 936
QY 301 ProTyrProMetAlaProProAsnGlnHisValSerProValGlnAlaLysTyrIleLeu 320
Db 937 CCTTATCCGATGGCGCCACCTAATGGCCACGTTTCACTGTGCAAGCCAAATATCATCTG 996
QY 321 LysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuLeuGlnLysIleLeuPro 340
Db 997 AAGACAGCTTCCATCTTTTCGAGACTGGCTATGAGCTTCTGCAAGTCACTTGCC 1056
QY 341 LeuLysSerPheThrAlaValCysGlnLysAspGlySerTyrPaspArgProMetProAla 360
Db 1057 CTGAAATCCCTTACTGCTGAGTGTGCAAGAAAGATGATCTTGGGACGGCCAAATGCCGCG 1116
QY 361 CysSerIleValAspCysGlyProProAspAspLeuProSerGlyLysArgValGluTyrIle 380
Db 1117 TGCAGCATTTGTTACTGTGGCTCTCTGATGATCTACCCAGTGGCGGAGTGAATCATC 1176
QY 381 ThrGlyProGlyValThrThrTyrLysAlaValIleGlnTyrSerCysGluGluThrPhe 400
Db 1177 ACAGTGTCTGAGTACACACCTCAAAAGCTGTGATTCAGTACAGCTGTGAAGAGACTTC 1236
QY 401 TyrThrMetLysValAsnAspGlyLysTyrValCysGluAlaAspGlyThrPheThrSer 420
Db 1237 TACACAAATGAATGATGATGATAATGTGTGTGTGAGGCTGATGATCTGTGACGAGC 1296
QY 421 SerLysGlyLysSerLeuProValCysGluProValCysGlyLeuSerAlaArgThr 440
Db 1297 TCCAAAGAGAAATAATCACTCCAGCTGTGAGCCCTGTGTGGACTATACAGCCGACAA 1356
QY 441 ThrGlyLysArgIleTyrGlyGlyGlnLysAlaLysProGlyLysPheProTrpGlnVal 460
Db 1357 ACAGGAGGCGGTATATATGAGGAGGCAAAAGGCAAAACCTGGTGATTTTCTTGGCAAGTC 1416
QY 461 LeuIleLeuGlyGlyThrThrAlaIleAlaGlyAlaLeuLeuTyrAspSerThrValLeuThr 480
Db 1417 CTGATATATGAGTGAACCAACAGCAGGAGCTTATATATGACAACTGAGGCTCTTAACA 1476
QY 481 AlaAlaHisAlaValTyrGluGlnLysHisAspAlaSerAlaLeuAspIleArgMetGly 500
Db 1477 GCTGCTCATGCCCTCATATAGCAAAACATGATGATTCGCGCTTGACATTTGGAATGGGC 1536
QY 501 ThrLeuLysArgLeuSerProHisTyrThrGlnAlaTrpSerGluAlaValPheIleHis 520
Db 1537 ACCCTGAAAAGACTATACACTCATATATACAAAGCTGTGTAAGCTGTTTATATACAT 1596
QY 521 GluGlyTyrThrHisAspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsn 540
Db 1597 GAAGTTATATCATATGATGCTGCTTGTGACATGACATGACATGATTAATTTGAATTAAC 1656
```

Qy	541	LysAlaValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGlnSer	560
Db	1657	AAAGCTGTAAATCATATGCAACATCACGGCTATTGTCTGCCAAGAAAGACTGTAAATCC	1716
Qy	561	PhenAlaArgThrAspAspIleGlyThrAlaSerGlyTrpGlyLeuThrGlnArgGlyPhe	580
Db	1717	TTTATGAGGACACAGATGACATGTGGAACCTGCATCTGCATGGGGATTAAACCCAAAGGGGTTTT	1776
Qy	581	LeuAlaArgAsnLeuMetCysValAspIleProIleValAspHisGlnLysCysThrAla	600
Db	1777	CTTCTCTGAAGATCTAAATGATGTCTGCACATACCGATGTGTGACATCAAAATAATGTACTGCT	1836
Qy	601	AlaTrpGlnLysProProArgProAlaArgGlySerValThrAlaAsnMetLeuCysAlaGly	620
Db	1837	GCATATGAAAGCCACCCCTTATCCAAAGGGAGAGTAACTGCTTAACATCTCTTGTGCTGGC	1896
Qy	621	LeuGlnSerGlyGlyLysAspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeu	640
Db	897	TTTAAAGAGTGGGGGCAAGCAGACTCTCAGAGGTGACAGCGGAGGGGCACTGGTGTCTTA	1956
Qy	641	AspSerGlnThrGluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsnCys	660
Db	1957	GATGTGTAACACAGAGAGGTGGTTTGGGGAGGAAATAGTCTCTGGGGTTCATGATATGT	2016
Qy	661	GlyGluAlaGlyGlnTrpGlyValTrpThrLysValIleAsnTrpIleProTrpIleGlu	680
Db	2017	GGGGAAACAGAGTCAGTATGAGAGCTTACACAAAGTTATTAACTAATATTCCTCGATCGAG	2076
Qy	681	AsnIleIleSerAspPhe	686
Db	2077	AACTAATATGATGATTTT	2094

```

RESULT 2
US-09-874-238-3
: Sequence 3, Application US/09874238
: Patent No. US20020082209A1
: GENERAL INFORMATION:
: APPLICANT: Jensenius, Jens Chr.
: APPLICANT: Thiel, Steffen
: TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
: TITLE OF INVENTION: USES FOR IT
: FILE REFERENCE: 09011-002003
: CURRENT APPLICATION NUMBER: US/09/874,238
: CURRENT FILING DATE: 2001-06-04
: PRIOR APPLICATION NUMBER: 09/054,218
: PRIOR FILING DATE: 1998-04-02
: APPLICATION NUMBER: 60/042,678
: FILING DATE: 1997-04-03
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 2475
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (37)...(2094)
US-09-874-238-3

Alignment Scores:
Pred. NO.:
0
Score: 3758.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
10
DB:
10
US-09-874-198-2 (1-686) x US-09-874-238-3 (1-2475)
Length: 2475
Matches: 686
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

QY 1 MetArgLeuLeuThrLeuLeuGlyLeuLeuCysGlySerValAlaThrProLeuGlyPro 20
 |||||
 Db 37 ATGAGCTGCTGCACCCCTCTGGGCTTCTGTGTGCTCGGTGGCCACCCCTTAGGCCG 96

Oy	21	LysTrpProGluProValPheGlyValGluAlaSerProGlyPheProGlyLysIleVal	40
Db	97	AAAGGGCTGAACCTGTGTTCGGGGCGCTGGCATCCCGGGCTTCCAGGGAGATATGCC	156
Oy	41	AsnAspGlnIuAgaTgTprThLeuThrAlaProProGlyTyrArgLeuArgLeuIyr	60
Db	157	AATAGACAGAGGGGGCTGGACCTGACTGACACCCCCGGCTACGGCTGGCCCTTAC	216
Oy	61	PheThrHisPheAspLeuGluLeuSerHisLeuGlySerIleAspPheValLysLeuSer	80
Db	217	TTACACCACTTCGACCTGGAGACTCTCCACACTCTGGCAGTACGACTTCTGCAAGCTGAC	276
Oy	81	SerGlyAlaLysValIleuAlaThrLeuGlyGlnGlnSerThrAspThrIuArgAla	100
Db	277	TGGGGGGCCAAAGGCTGGCGCCAGCTGTGGGGGAGAGACACACAGGAGGGGCC	336
Oy	101	ProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleThrPheArgSerAsp	120
Db	337	CTTGGAAGAAGACACTTTCATCGCTGGGTGGCTCCAGCTGGACATTATCTTCGGCTCCGAC	396
Oy	121	TyrSerAsnGlyLysProPheThrGlyPheGluAlaPheTyrAlaAlaGluAspIleAsp	140
Db	397	TATCTCCAAAGGAAGCCGTTCCAGGGGTTTCGAGGCTTTATGCAAGCCAGACATTTGAC	456
Oy	141	GluGlyGlnValAlaProGlyGluAlaProThrCysAspHisHisCysHisHisHisLeu	160
Db	457	GAGTGCCAGATGGGGCCCCGGAGAGGGCCCCACCTGGGACACCACTGGCACCAACCACTTG	516
Oy	161	GlyGlyPheTyrCysSerCysArgAlaGlyTyrValIleuHisArgAsnLysAspThrCys	180
Db	517	GGCGGGTTTCTACTCTCTCTCCCGGACGGCTACGCTCTGACCCGTAAACAGGGCACTGGC	576
Oy	181	SerIleLeuGlySerGlyGlnValPheThrGlnArgSerGlyGlnLeuSerSerProGlu	200
Db	577	TTGACCCCTGTGCTCCGGCCAGGTCTTCCACCAAGAGGTCTGGGAGCTCAGCAGCCCTGAA	636
Oy	201	TyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGlnGly	220
Db	637	TACCACGGCGCGTATCCCAAACTCTCCAGTTGCATTCACGATCAGCATCAGCTCGGAGGAGGG	696
Oy	221	PheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeu	240
Db	697	TTTCAGTGTCAATTCGGACTTTGTGGATCTCTTCGATGTGGAGACACACCTTAAACCTCTG	756
Oy	241	CysProTyrAspPheLeuLysIleGlnThrAspArgGlnGlnHisGlyProPheGlyGly	260
Db	757	TGTCCCTACGACTTTCACAGATTCAACACGACAGAGAAACATGGCCCATTTCTGTGGG	816
Oy	261	LysThrLeuProHisArgIleGluThrLysSerAspThrValThrIleThrPheValThr	280
Db	817	AAGACATTTGGCCCCACAGATTGAACAAAAGCAACAGCGTGACATCCATTTGTGCACA	876
Oy	281	AspGluSerGlyAspHisThrProLysIleHisTyrThrSerThrAlaLeuProCys	300
Db	877	GATGAATTCAGAGACCCACACAGGCTGGAAATCTCACTACACAGACACACCGAGCGTTTGC	936
Oy	301	ProTyrProMetAlaProProAsnGlnHisValSerProValGlnAlaLysTyrIleLeu	320
Db	937	CGTTATCGAGTGGGGCCACTAATGGCCAGCGTTTCACTGCTGCMAAGCCAAATACATCTTG	996
Oy	321	LysAspSerPheSerIlePheCysGlyThrGlyTyrGluLeuGlnGlnLysLeuPro	340
Db	997	AAAGACAGCTTCTCCACTTTTTCGAGACAGGCTAAGACTTCTGCAAGCTCACTTGGCC	1056
Oy	341	LeuLysSerPheThrAlaValLysGlnLysAspGlySerThrAspArgProPheProAla	360
Db	1057	CTGAATATCTTTACTGAGTTTGTCAAGAAAGATGGATTTGGGGACCGGCAATGCCCGCG	1116
Oy	361	CysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluIuThrIle	380
Db	1117	TGCAGCATTTGTGACTGTGGCCCTCTCGATGATCTTCCAGTGGCCGATGGAGTACATTC	1176

337	CTGGCAGGACGACATTCTCTACGCTGGGCTCCAGGCTGACATTACCTTCGGCTCCGAC	396
QY	121 TyfSerAsnGluLysProPheThrGlyPheGluAlaPheThrAlaIleAspLysPheAsp	140
Db	397 TACTCCAAACGAGAAAGCCGTTCCAGGGGTTCCAGACCTTTATGCAACCGACGACATTGAC	456
QY	141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisIleCysHisAsnHisLeu	160
Db	457 GAGTGCAGGAGGGCCCGGGAGAGAGCGCCACCTGCGACACCACTGGCCAAACCACTGG	516
QY	161 GlyGlyPheThrCysSerCysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCys	180
Db	517 GGGCGTTTCTACTGCTCTCTCCCGGAGGTACGTCTCCACCGTAACAGGGCACTGCG	576
QY	181 SerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGluLeuSerSerProGlu	200
Db	577 TCAGCCCTGCTCTCCGGCCAGGTCTTCACCCAGAGGTCTGGGAGTCAAGCACTCTGAA	636
QY	201 TyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSerLeuGluGluGly	220
Db	637 TACCACGCGCGGTATCCCAAACTCTCCAGTTGCATTCACGCACTACGCTGGAGGAGGG	696
QY	221 PheSerValIleLeuAspPheValGluSerPheAspValGluThrHisProGluThrLeu	240
Db	697 TTTCAGTGCATTCCTGGCATTTGTGTGAGTCTTCGATGTGAGAGCAACCTCTGAAACCTG	756
QY	241 CysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCysGly	260
Db	757 TGTCCCTTACGACATTCTTCAAGATTCAAAACGACAGAGAAACATGGCCCATTCGTGTGG	816
QY	261 LysThrLeuProHisArgIleGluThrLysSerAspThrValThrIlePhePheValThr	280
Db	817 AAGACATTGCCCCCAAGAGATTGAACAAAAGCAACCGTGACCTACCTTTGTTCACA	876
QY	281 AspLeuSerGlyAspHisThrGlyTyrLysIleHisTyrThrSerThrAlaGlnProCys	300
Db	877 GATGAAATCAGAGACACACACAGCGTGGAAAGATCCACTACACAGCAAGCCGACGCTTGC	936
QY	301 ProTyrProMetAlaProProAsnGlyHisValSerProValGlnAlaLysTyrIleLeu	320
Db	937 CCTTATCCGATGGGGCCACCTTAATGGCCACGTTTCACTGCTCAAGCCAAATATCATCTCTG	996
QY	321 LysAspSerPheSerIlePheCysGluThrGlyTyrGlyLeuLeuGlnGlyHisLeuPro	340
Db	997 AAAAGACAGCTTCTCCATCTTTTCGAGACAGCTGGAAGACCTTCTCCAAAGCTACTTGGCC	1056
QY	341 LeuLysSerPheThrAlaValLysGlnLysAspGlySerTyrAspArgProMetProAla	360
Db	1057 CTGAATTCCTTTACTGCGAGTTTGTCAAGAAAGATGATCTTTGGGACCGCGCAATGCCCGC	1116
QY	361 CysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgValGluTyrIle	380
Db	1117 TGCAGCATTTGTGACTGTGGCCCTCTCTGATGATCTTCCAGTGGCCGAGTGGAGTACATTC	1176


```

Db      706 CCCCCTGACCTGGCGCTGCAACCTACAGCATCCGGGTGGAGCGGGCCCTCACCCCTGCACCTC 765
Qy      226 AspPheValIuSerPheAspValIGluThrHisProGluThrLeuGlyProTyrrAspPhe 245
Db      766 AAGTCTGGAGCTTTGATTGATGATGACACCAAGCAATACATACCTCCCTATGAGCAG 825
Qy      246 LeuIleuIleGlnThrAspArgGluGlnHisGlyProPheCysGlyLysThrLeuProHis 265
Db      826 CTACAGATCATATGCCAAGGAGAACATGTGGAGATCTCTGTGGAGAACAAAGCCCCC 885
Qy      266 ArgIleGluThrLysSerAsnThrValThrIleThrPheValIleThrAspGluSerGlyAsp 285
Db      886 GACCTGACACCAAGCAAGCAATGCTGATCTGCTGTTCTTACAGATGATGATGGGGGAC 945
Qy      286 HisThrGlyThrLysIleHisIleThrSerThrAlaGlnProCysProTyrrPrometAla 305
Db      946 AGCCGGGCTGGAAAGCTGCGCTACACCCAGAGATCATCAAGTGCCTCCCAAGACC 1005
Qy      306 ProProAsn---GlnHisValSerProValGlnAlaLysThrIleLeuLysAspSerPhe 324
Db      1006 CTAGAGAGATTACACATCATCCAGAACCTGCACCTCAGTACAGTCCCTGACTACTTC 1065
Qy      325 SerIlePheCysGluThrGlyThrGlyLeuLeuGlnGlnHisLeuProLeuLysSerPhe 344
Db      1066 ATTGCTACCTGCAGCAAGCAAGCTACAGCTCATAGAGGGAGAACAGGAGTGTGATCTTC 1125
Qy      345 ThrAlaValIleGlnLysAspGlySerThrPaspArgPrometProAlaCysSerIleVal 364
Db      1126 ACAGCTGTGTGCCAGAGATGAGTGGACGTGCTGCCATCTGCCAGATCCAAAGATACG 1185
Qy      365 AspCysGlyProProAspAspLeuProSerGlyArgValIleValIleThrGlyProGly 384
Db      1186 GACTGTGGGAGGCCCGGAACCTGCCTAATGATGACTCCGTTACACACACACATGGGA 1245
Qy      385 ValThrThrTyrlsAlaValIleGlnTyrrSerCysGluIleThrPheTyrrMetLys 404
Db      1246 GTGAACACCTACAAAGCCCTATCTCAGTACTGATGAGCAGATATACAGATGACAG 1305
Qy      405 Val-----AsnAspGlyLysTyrrValIleGlnLysAlaAspGlyPhe 417
Db      1306 ACCAGAGCTGGCAGCAGGAGGTGTGACCAAGGGGTGTACCTGCACAGCAGACAGGCGATT 1365
Qy      418 Trp---ThrsSerLysGlyGluLysSerLeuProValIleGlyLysIleGlyGly--- 435
Db      1366 TGGAGAGATTAACAGAAAGGAGAGAGAG---ATTCTCTGGTGTCTTCCAGTGTGTGGAGAG 1422
Qy      436 ---LeuSerAlaArgThrThrGlyGlyArgIleTyrrGlyGlnLysAlaLysProGly 454
Db      1423 CCGGTGAACCCCGTGAACAGAGGAGCGCATATCGGAGCGCAAAAAGCCAAAGATGGCG 1482
Qy      455 AspPheProFrpGlnValIleuIleLeuGlyGlyThrThrAlaAlaGlyAlaLeuLeuTyrr 474
Db      1483 AACTTCCCTGGGAGGTTTCACCAACATCCACGGCGCGGGCGCGCGCTGCTGGCG 1542
Qy      475 AspAspThrValLeuThrAlaAlaHisAlaValTyrrGluGlnLysHisAspAla----- 492
Db      1543 GACCGGTGATCTCTACAGCTGCCACACCTGTATCCCAAGAACAGAGCGCAAGC 1602
Qy      493 ---SerAlaLeuAspIleArgMetGly-----ThrLeuLysArgLeuSer 506
Db      1603 AACGGCTCTTGGATGTCTGCGGCACACAAATGTGGAAAGAGCTCATGAAAGCTAGGA 1662
Qy      507 ProHisTyrrThrGlnIleArgSerGluAlaValAlaPheIleHisGluGlyTyrrThrHisAsp 526
Db      1663 AATCACCCCATCCGC-----AGGGTACGCTTCCACCCGCGAGTACCGCTCAGAT 1710
Qy      527 Ala-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValIle 544
Db      1711 GAGTCTACAAATTTTGGAGGGGAGACATCGCCCTGCTGTGAGAGTGGAAATATATGACCCCTG 1770
Qy      545 AsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThr 564

```

```

Db      1771 GGTCCCAACCTCCCTCCCATCTGCTCCCT-----GACAACGATACCTTCTAGACCTG 1824
Qy      565 AspAspIleGlyThrAlaSerGlyTrpGlyLeuThrGlnArgGlyPheLeuAlaArgAsn 584
Db      1825 GCCTTGATGGCTATGTCACATGCTTGGGGTCAAGGAGAGAG---ATTGCTCATGAC 1881
Qy      585 LeuMetIleValAspIleProIleValAlaSerHisGlnLysCysThrAlaAlaIleTyrrGluLys 604
Db      1882 CTCAGGTTTCTCCGTCGCGCCGTAAGCATATCCACAGGCCGT----- 1923
Qy      605 ProProTyrrProArgLysSer-----ValThrAlaAsnMetLeuCysAla 619
Db      1924 GAGAAGTGGCTCCGGGGAAGAATAGAGATGATGTCTTCTCAAAAACATGTTCTGTGCT 1983
Qy      620 GlyLeuGlnSerGlyGlyLysAspSerCysArgGlyLysPheSerGlyGlyAlaLeuValPhe 639
Db      1984 GGACACCATCTCTTAAGCAGAGCGCTGCCAGGGGATGTGGGGCGTTTTCAGTA 2043
Qy      640 LeuAspSerGluThrGluArgTrpPheValIleGlyTyrIleValSerTrpGlySerMetAsn 659
Db      2044 AGGGAGCCCGAACAACATGATCGCTGGGTGGCCAGCGGATGCTGTCTGTGGGC---ATCGGG 2100
Qy      660 CysGlyGluAlaGlyGlnTyrrGlyValTyrrThrLysValIleAsnTyrrIleProTrpIle 679
Db      2101 TGCAGCAGGGGC-----TATGCTTCTTACACCAAGTCTCAACTPACGTGAGCTGGATC 2154
Qy      680 Glu 680
Db      2155 AAG 2157

RESULT 4
US-10-044-090-613
; Sequence 613, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 613
; LENGTH: 2555
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1003386CB1
US-10-044-090-613

Alignment Scores:
Pred. No.: 8,25e-130 Length: 2555
Score: 1412.00 Matches: 301
Percent Similarity: 56.73% Conservative: 108
Best Local Similarity: 41.75% Mismatches: 248
Query Match: 37,578 Indels: 64
DB: 12 Gaps: 22

US-09-874-198-2 (1-686) x US-10-044-090-613 (1-2555)
Qy      1 MetArgLeuLeuThrLeuLeu-----GlyLeuLeuCys-----GlySerValAla 15
Db      120 ATGTGCTCTTGATCCCTCGGTGGCCGCGCTTCTTCTGAGGGCAGAGAGGCTCCATT--- 176
Qy      16 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 35
Db      177 ---CCCATC-----CCTCAAGAGTTATTTGGGAGAGTGACTTCCTTCGTGTTTC 221
Qy      36 ProGlyGluTyrrAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyrr 55
Db      222 CCCAAGCTTACCCCAACAACATTGTAAGAACCAACCACTGTGATCACAGTCCCGCCAGGGATAC 281

```

QY	56	ArgLeuAtrGLeuTYrPheIThrHisPheAspLeuGlnLeuSerHisLeuGlyGluTYrAsp	75
Db	282	AGGGGAAGCTCTGCTTCCACAGACTTGACCTGGACCTCTGAAGCTGCTTATGAT	341
QY	76	PheValIysLeuSerSerGlyAlaIysValLeuAlaThrLeuGlyGlnLeuSerThr	95
Db	342	TATGTCAAGATCTGTGGTATAGAAAGCTGGGAGGATGTGTGGGCACTGGGTTCT	401
QY	96	AspThrGluAlaProGlyIysAspThrPheTYrSerLeuGlySerSerLeuAspHis	115
Db	402	CCACTGGGCAACCCCGGGGAAGAAAGAAATTTATGTCCCAAGGAACAAAGATGCTGTG	461
QY	116	ThrPheArgSerAspTYrSerAsnGlyPro-----PheThrGlyPhe	130
Db	462	ACCTTCACACAGACTTCTCCACAGAGAAATGGACATCATGTCTTACAAAGGCTTC	521
QY	131	GluAlaPheTYrAlaAlaGluAspIleAsnGlyCys-----GlnValAlaProGlyIu	148
Db	522	CTGGGCTACTACCAAGCTGTGAGACTTGATGTGCTCCCGGAGCAAAATCAGGGAG	581
QY	149	-----AlaProThrCysAspHisHisCysHisAsnHisLeuGlyIlePheTYrCys	165
Db	582	GAGGATCCCAAGCCCAAGTCGACACACTGTGTACAACTACCTGTGAGAGCTACTTCTGT	641
QY	166	SerCysArgAlaGluTYrValIleuHisArgAluAspThrCysSerAlaLeuGlySer	185
Db	642	TCTCTCCCTCCAGGCTTCAAGCTTCACAGAAACAGAGATTCCTGCAAGCTAGTACAC	701
QY	186	GlyAluValPheThrGlnAlaGlySerGlyIleuSerSerProGlyIuTYrProArgProTYr	205
Db	702	AGCAGAGCTGACAGGAGGATCTGAGTACATCTCCAGCTGGAGTACCTGGTCTTAC	761
QY	206	ProIuysLeuSerSerCysTYrThrSerTleSerLeuGlnGluGlyPheSerValIleLeu	225
Db	762	CCCCCTGCACCTGCCTCACTACAGATCGGGGTGGAGGGGGGCTCACTGCACATC	821
QY	226	AspPheValIuSerPheAspValGlnThrHisProGlnThrLeuGlyProTYrAspPhe	245
Db	822	AAGTTCCTGGAGCCTTTTGATTTGATGAGCCACACAGATACATCACTGCCCCATTAACAG	881
QY	246	LeuIysIleGlnThrAspArgGlnGlnHisGlyProPheCysGlyIuTYrThrLeuProHis	265
Db	882	CTACAGACTTATGCCAAGGGAAGAACATTTGGGAGGATGTGTGGGAAGCAAGGCCCC	941
QY	266	ArgIleGluThrIysSerAsnThrValThrIleThrPheValIThrAspIuSerGlyAsp	285
Db	942	GACCTCGACACACACAACTGCTGTGATCTCTGTCTTCCACAGATGATGCTGGGGCAC	1001
QY	286	HisThrGlyTrpIysIleHisTYrThrSerThrAlaGlnProCysProTYrProMetAla	305
Db	1002	AAGCGGGGCTGGAAAGCTGGCCCTACACCCAGACATCATCAAGTGGCCCAAGCCCAAGAC	1061
QY	306	ProProAsn---GlnHisValSerProValGlnAlaIuTYrIleLeuIysAspSerPhe	324
Db	1062	CTAGACGAGTTCCACCATCATCAAGAACCTGCACCTCCAGTACAGATTCGCTGACTACTTC	1122
QY	325	SerIlePheCysGluThrGlyTYrGlyLeuLeuGlnGlyHisLeuProLeuIysSerPhe	344
Db	1122	ATTGCTACTCTCAGACCAAGCTCACTACAGCTCATAGAGGGAACACAGGTGCTGATTCCTTC	1181
QY	345	ThrAlaValIysGlnIuysAspGlySerTrpAspArgProMetProAlaCysSerIleVal	364
Db	1182	ACAGCTGTCTCCCGAGATGTAGTGCACCTGGCACTGTCCACCGAGTCAAGATTCAG	1241
QY	365	AspCysGlyProProAspAspLeuProSerGlyArgValGluIuTYrIleThrGlyProGly	384
Db	1242	GACTGTGGGACGCCCAAACTCTCCATAATGGTGACTCCGTTTACCAACCAATGGGA	1301
QY	385	ValThrTYrTYrIysAlaValIleGlnTYrSerCysGluGluIuThrPheTYrThiMetIys	404
Db	1302	GTGAACACCTTACAGGCCCGTATTCACAGTACTACTGCCATGAGGCATATTACAAAGTGCAG	1361
QY	405	Val-----AsnAspGlyIuTYrValCysGlnAlaAspGlyPhe	417

[illegible]

```

: APPLICANT: Mishra, Vishnu
: APPLICANT: Mezes, Peter S
: APPLICANT: MacDougall, John
: TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 15966-697 CIP
: CURRENT APPLICATION NUMBER: US/09/808,602
: CURRENT FILING DATE: 2001-03-14
: PRIOR APPLICATION NUMBER: 09/800,198
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: 60/186,596
: PRIOR FILING DATE: 2000-03-03
: NUMBER OF SEQ ID NOS: 114
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO: 92
: LENGTH: 2386
: TYPE: DNA
: ORGANISM: Homo sapiens
: 808-602-92

```

Alignment Scores:

```

Pident. No.: 1,17e-129 Length: 2386
Score: 1410.00 Matches: 301
Percent Similarity: 56.73% Conservative: 108
Best Local Similarity: 41.75% Mismatches: 248
Query Match: 37.52% Indels: 64
DB: 9 Gaps: 22

```

US-09-874-198-2 (1-686) x US-09-808-602-92 (1-2386)

```

QY 1 MetArlLeuLeuThrLeuLeu-----GlyLeuLeuLeuCys-----GlySerValAla 15
DB 52 ATGTGGCTCTTGTACCTCGTGGTCCGCGCCCTCTTGTGCGGCGAGGCGCTCCATT--- 108
QY 16 ThrProLeuGlyProLysTrpProGluProValPheGlyArlGlyLeuAlaSerProGlyPhe 35
DB 109 ---CCATC-----CCTCAGAGTTATTGGGAGGTGACTTCCCTGTCTTC 153
QY 36 ProGlyGluArlAlaAsnAspGlnGluArlGlyArlTrpThrLeuThraLapProGlyTyr 55
DB 154 CCAGAGCTTACCCCAACACATTGGAACACACAGCTGTATCATACAGTCCCGGAGGATAC 213
QY 56 AylLeuArlGlyTrpThrTrhAspLeuGlyLeuSerHisLeuCysGlyTyrAsp 75
DB 214 AGGTTAAGTCTCTCTCGACGAGTTTGTGAGCTTGTGAAGCTGCTTCTATGAT 273
QY 76 PheValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuGlyGlnGluSerThr 95
DB 274 TATGTCAAGATCTCTGCTGTAAAGAAAGCTTGGGAGTTCGTGTGGACACTGGTCT 333
QY 96 AspTrhGluArlArlProGlyLysAspTrhPheTyrSerLeuGlySerSerLeuAspIle 115
DB 334 CCACGTGGGCAACCCCGGGAAGAAGATTATTATGCCAAGGAAACAGATGCTGCTG 393
QY 116 ThrPheArlSerArlTrpTrhAsnGlnGluLysPro-----PheTrhGlyPhe 130
DB 394 ACCTTCCACACAGATCTTCCACAGAGGAATGGACATCATGTTCTTCAAGGGCTTC 453
QY 131 GluAlaPheTyrAlaLagLysAspIleAspGlyCys-----GlnValAlaProGlyGlu 148
DB 454 CTGGCTTACCAACAGCTGTGGACCTTGATGAATGCTTCCCGGAGCAAAATTAGGGGAG 513
QY 149 -----AlaProTrhCysAspHisHisCysHisAsnHisLeuGlyGlyPheTyrCys 165
DB 514 GAGAGTCCCGACAGCCAGTGGCCAGACACTGTGTACAACTAGCTTGGAGCTACTTCTGT 573
QY 166 SerCysArlGlyArlGlyTrpValLeuHisArlGlyArlGlyTrpCysSerAlaLeuCysSer 185
DB 574 TCTGTGCGTCCAGGCTATAGCTTCAAGAGACAGGACATTCGCGCAGGCTGAGTGAGC 633
QY 186 GlyGlnValPheTrhGlnArlGlySerGlyGluLeuSerSerProGlyTyrProArlProTyr 205
DB 634 AGGAGACTGTACACGAGGAGCATGAGGCTACATCTCCAGCTGGAAGTACCTCGTCTTAC 693

```

```

QY 206 ProLysLeuSerSerCysTrhTrpSerIleSerLeuGlnGluGlyPheSerValIleLeu 225
DB 694 CCCCTGACCTGCGGTGCAACATACACATCCGGTGGAGGGGGGCTCCAGTGCACCTC 753
QY 226 AspPheValGluSerPheAspValGluTrhHisProGluTrhLeuGlyProTyrAspPhe 245
DB 754 AAGTCTGTGAGCCCTTTGTATTGATGACACCACCAATACAGTCCGCTTATGACAC 813
QY 246 LeuLysIleGlnTrhAspArlGlnGluHisGlyProPheCysGlyLysTrhLeuProHis 265
DB 814 CTACAGATCTATGCAACAGGGAAGAACATTGGGAGTCTGTGGAGAACAGAGCCGCC 873
QY 266 AylGluLeuTrhLysSerAsnTrhValTrhIleTrhPheValTrhAspGluSerGlyAsp 285
DB 874 GACCTGACACACAGACAGCAATGCTGTGATCTGCTGTTCACAGATGATGCTGGGGAG 933
QY 286 HisTrhGlyTrpLysIleHisTrhSerTrhAlaGlnProCysProTyrProMetAla 305
DB 934 AGCCGGGGCTGGAAGCTGACCTACACCACAGATCATCAAGTGCCTCCAGCCAGACC 993
QY 306 ProProAsn---GlyHisValSerProValGlnAlaLysTrpIleLeuLysAspSerPhe 324
DB 994 CTAGACAGATTCACCATCATCCAGAACCTGACCTCAGTACCATGTTCCGTGACTTCTC 1053
QY 325 SerIlePheCysGluTrhGlyTrpGluLeuGlnGlyHisLeuProLeuLysSerPhe 344
DB 1054 ATTGCTACATCTGCAAGCAAGCTGACATCATAGAGGGGAACAGAGTTCGATTCCTTC 1113
QY 345 ThrAlaValCysGlnLysAspGlySerTrpAspArlProMetProAlaCysSerIleVal 364
DB 1114 ACAGCTGTCTGCGCAGATGATGTCAGCTGTCATGTCCTGATGCCAGATCAGATCAG 1173
QY 365 AspCysGlyProProAspAspIleProSerGlyArlValGluTrpIleTrhGlyProGly 384
DB 1174 GACTGTGGCAGCCCGAAGCTGCTTAATGTGACTTCCGTTTACACACACCAATGGGA 1233
QY 385 ValTrhTrpTyrLysAlaValIleGlnTyrSerCysGlnGluTrhPheTrhMetLys 404
DB 1234 GTGACACACCTACAAAGGCCGTATCCAGATGATGATGATGATGATGATGATGATGAT 1293
QY 405 Val-----AsnAspGlyLysTrpValLysGlnAlaLysAspGlyPhe 417
DB 1294 ACCAGAGCTGCGACAGGAGTCTGAGCAAGGGGTGACCTGACACGACGACAGGGCAT 1353
QY 418 Trp---ThrSerSerLysGlyGluLysSerLeuProValLysGlyProValLysGly 435
DB 1354 TGGAGATGATGAAGAGAGAGAGAG---ATTCTCTGCTGCTTCCAGTGTGTGGAG 1410
QY 436 ---LeuSerAlaArlTrhTrhGlyArlGlyIleTyrGlyGlyGlnLysAlaLysProGly 454
DB 1411 CCCGTGAACCCCTGMAACAGAGGACAGCGCATCATGCGAGGCAAAAGCCAGATGGGC 1470
QY 455 AspPheProTrpGlnValLeuIleLeuGlyGlyTrhAlaAlaGlyAlaLeuLeuTyr 474
DB 1471 AACTTCCCGCGAGGCTTCCACCAACATCCAGGGCGCGGGCGGCGCTGTGCGGC 1530
QY 475 AspAsnTrpValLeuTrhAlaAlaHisAlaValTrpClnGlnLysHisAspAla----- 492
DB 1531 GACCGGTGATCTTCACAGCTGCCACACCTGTATTCACAGGAACAGAGGCAAGC 1590
QY 493 ---SerAlaLeuAspIleArlGlyMetGly-----ThrLeuLysArlGlySer 506
DB 1591 AAGCGCTCTTGGATGTGCTTGGGCGACACAAGTGTGAAGAGCATGACATGATGAGA 1650
QY 507 ProHisTrpTrhClnAlaTrpSerGlnAlaValPheIleHisGluGlyTrpTrhHisAsp 526
DB 1651 AATCACCCCATCCG-----AGGCTCAGGCTCCACCCGAGTACCTCAGAGAT 1698
QY 527 Ala-----GlyPheAspAsnAspIleAlaLeuLysLeuAsnLysLysValIle 544
DB 1699 GAGTCTACAAATTTTGAAGGGGACATGCGCTGTGAGCTGGAATAATGATGACACCTG 1758
QY 545 AsnSerAsnIleTrhProIleCysLeuProArlGlyGlnAlaGluSerPheMetArlTrh 564

```


[illegible]

Db	1940	GTTCGCCCTCAAGGGGCGACAGGTTACTGTAAGTCCTTTAAGAAATGCAAAACAACTGMAA	1999			
Qy	602	TyrGluLysProPrrTyrrProArgGlySer-----ValThrAlaIasnMetLeuCysAla	619			
Db	2000	GTGGAGAAACCCACAGCAGATGACAGAGGCGCTATGTTTCATCTCTTAACATGATCTGTGCT	2058			
Qy	620	GlyLeuGluSerGlyGlyLysAspSerCysArgGlyLysSerGlyLysAlaLeuValPhe	639			
Db	2060	GGAGAGAGAGAG--GGCATGAGTACCTGTAAAGGGGACAGTGGGGGCTTTCGTGA	2118			
Qy	640	LeuAsp---SerGluTrpGluArgTrpPheValGlyIleValSerTrpGlySerMet	658			
Db	2117	CAGATGCCCATGACACAGACCAATTCATACGACAGCTGGCGCTGTCTGTCTCGGGGGCC---	2173			
Qy	659	AsnCysGlyLysAlaGlyGlnTyrrGlyValTyrrThrLysValIleAsnTyrrIleProTrp	678			
Db	2174	-----CAGTGTGGGACCTTATGGGCTCTACACACAGGGGTAAAGAACTATGTTGACTGG	2224			
Qy	679	Ile 679				
Db	2225	ATA 2227				
RESULT 8						
US-09-925-301-182						
; Sequence 182, Application US/09925301						
; Patent No. US20020052308A1						
; GENERAL INFORMATION:						
; APPLICANT: Rosen et al.						
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies						
; FILE REFERENCE: PA106						
; CURRENT APPLICATION NUMBER: US/09/925,301						
; PRIOR FILING DATE: 2001-08-10						
; PRIOR APPLICATION NUMBER: PCT/US00/05882						
; PRIOR FILING DATE: 2000-03-08						
; PRIOR APPLICATION NUMBER: 60/124,270						
; PRIOR FILING DATE: 1999-03-12						
; NUMBER OF SEQ ID NOS: 1694						
; SOFTWARE: PatentIn Ver. 2.0						
; SEQ ID NO 182						
; LENGTH: 2725						
; TYPE: DNA						
; ORGANISM: Homo sapiens						
; FEATURE:						
; NAME/KEY: misc_feature						
; LOCATION: (2713)						
; OTHER INFORMATION: n equals a,t,g, or c						
US-09-925-301-182						
Alignment Scores:						
Pred. NO.:	1,41e-112	Length:	2725			
Score:	1238.00	Matches:	268			
Percent Similarity:	52.78%	Conservative:	102			
Best Local Similarity:	38.23%	Mismatches:	283			
Query Match:	32.94%	Indels:	48			
DB:	10	Gaps:	18			
US-09-874-198-2 (1-686) x US-09-925-301-182 (1-2725)						
Qy	1	MetArgLeuLeuThrIleLeuLeuGlyLeuLeuCysGlyIleValAlaIleThrProLeuGlyPro	20			
Db	220	ATGGTGTCATGCTGCTGTTTCATCTTTTGGCATGGGTTTATGCTGAGGCT-----	270			
Qy	21	LysTrpProGluProValIlePheGlyArgLeuAlaSerProGlyIlePheProGlyIleTyrrAla	40			
Db	271	-----ACCATGTAATGGGAGAGATCTCTCCCTCAACTATCTTCAGGCAATATCCC	318			
Qy	41	AsnAspGlnGluArgArgTrpThrIleLeuThrAlaProProGlyIleTyrrArgLeuArgLeuTrp	60			
Db	319	AGTGAGGAGAGAAATCTTGAGCATAGAAGTCTCGAAGGGGTATGGGATTCACCTTAC	378			
Qy	61	PheThrHisPheAspLeuGluLeuSerHisLeuCysGlyIleTyrrAspPheValLysLeuSer	80			
Db	379	TTTCACCATCTGGCAATTTGCTGTCTAGAGAACTGTGCGTATAGACTCAGTGCAGATATATC	438			

QY 81 SerGlyAlaLysValLeuAlaThrLeuCysGlyGlnGluSerThrAspThrGluArgAla 100
 |||||
 Db 439 TCAGAGACACTGAAAGAGGAGGCTCTGGACAGAGGAGCATCAATCCCACTCT 498
 QY 101 ProGlyLysAspThrPheThrLeuGlySerSerLeuAspIleThrPheArgSerAsp 120
 |||||
 Db 499 CCATGTTGGAGAGCTTCCAGAGTCCCATCAACCAATCCAGTGTCTTAAAGTGGAC 558
 QY 121 TyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaIleGluAspIleAsp 140
 |||||
 Db 559 TTTTCCAAATGAAAGCGCTTTACGGGGTTTGCTGCATACACTTCCACAGACATAAAT 618
 QY 141 GluCysGlnValAlaProGlyGluAlaProThrCysAspHisHisAsnHisLeu 160
 |||||
 Db 619 GAATGCACAGATTTGTGA---GATGTCCCT---GTAGCCACTTTCGCAACATTTTCATT 672
 QY 161 GlyGlyPheThrCysSerCysAlaGlyTyrValLeuHisArgAsnLysArgThrCys 180
 |||||
 Db 673 GGTGGTACTTCTGCTCTGCCCCCGGGAATTTCTCATGATGACATGAAGAAATTCG 732
 QY 181 SerAlaLeuSerSerGlyGlnValPheThrGluArgSerGlyGluLeuSerSerProGlu 200
 |||||
 Db 733 GGAGTTAATTCGAGTGGGATGTATTCACCTGCACCTATTGGGAGATTGCAGATCCCAAT 792
 QY 201 TyrProArgProThrProLysLeuSerSerCysThrTyrSerIleSerLeuGlnGly 220
 |||||
 Db 793 TATCCCAACATATCCAGAGAACTCAAGGTGTGAATACAGATCCGGTGGAGAAAGG 852
 QY 221 PheSerValIleLeuAspPhe---ValGluSerPheAspValGluThrHisProGluThr 239
 |||||
 Db 853 TTCACAGTGGTGGTACCTTGGCGAGAGAAAGATTTGATGTGAACACAGCTGACTCAGCG 912
 QY 240 LeuCysProTyrAspPheLeuLysIleGlnThrAspArgGluGlnHisGlyProPheCys 259
 |||||
 Db 913 GGAACCTGCCCTTCACAGTTAGTTTGTTCGACGAGATCGCAATTTGCTTACTGT 972
 QY 260 GlyLysThrLeuProHis-----ArgIleGluThrLysSerAsnThrValThrIleThr 277
 |||||
 Db 973 GGTCAATGATTCCTCGGSCCTTAATATTGAACCAAGATGCTGTGATATCANTC 1032
 QY 278 PheValThrAspLysSerGlyAspHisThrGlyTyrPheLysIleHisThrIleThrAla 297
 |||||
 Db 1033 TTCACAACTGATCTTACAGAGGCAAAAAAGGCTTGGAATCTGCATCATGAGATCCCA 1092
 QY 298 GlnProCysProTyrProMetAlaProAsnGlyHisValSerProValGlnAlaLys 317
 |||||
 Db 1093 ATGCCCTGCCCT---AAGGAAGACACTCCCAATTTCTGTTGGAGCCTCGGAAGGCCAAAA 1149
 QY 318 TyrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGluLeuGlnGly 337
 |||||
 Db 1150 TATGCTTTAGAGATGTGTGTCAGATACCTCTCTGATGGTGTGAAGTTGTGAGGGA 1209
 QY 338 HisLeuProLeuLysSerPheThrAlaValCysGlnLysAspLysSerThrPheAspArgPro 357
 |||||
 Db 1210 CGTGTGGTGCAACATCTTCTATTCGACTTGTCAAGCAAGTGAAGTGAATTCCTC 1269
 QY 358 MetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGlyArgVal 377
 |||||
 Db 1270 AAACCTAAATGTCACCTGTGGACTGTGCATTCCTGAATCCATTGAGATGGTTAAAGTT 1329
 QY 378 GluTyrIleThrGlyProGlyValThrThrTyrLysAlaValIleGlnThrSerCysGlu 397
 |||||
 Db 1330 GAA-----GACCCAGAGACACTTGTGTGTTCTGTCTATCCGATCAACTGTGAG 1380
 QY 398 GluThrPheThrThrMetLys---ValAsnAspGlyLysTyrValCysGluAlaAspGly 416
 |||||
 Db 1381 GACCCATATTATCTACTGAGAAATGAGAGAGGTGGAGAGATCTACCTGCTGTGTAACGGG 1440
 QY 417 PheThrPheThrSerSerLysGlyLysSerLeuProValCysGluProValCysGlyLeu 436
 |||||
 Db 1441 AGCTGGGTGAATAGGTGCTGGGCCCGGAGCTGCCAATATGTGTTCACAGTCTGTGAGTTC 1500

QY 437 SerAlaArg-----ThrThrGlyGlyArgIleTyrGlyGlyGlnLysAlaLysProGly 454
 |||||
 Db 1501 CCAGAGAACCCCTTTGAGAAAAAACAGAGGATTAATTGGAGATCCGATGCAATATTATA 1560
 QY 455 AspPheProThrGlnValLeuIleLeuGlyGlyThrThrAlaIleGlyAlaLeuLeuTyr 474
 |||||
 Db 1561 AACCTCCCTGGCACTG---TTCCTTGAACAAACCCATGGGCTGTGGACCGCTCAATTAAT 1617
 QY 475 AspAsnThrValLeuThrAlaIleHisAlaValTyrGluGlnHis----- 490
 |||||
 Db 1618 GAGTACTGGGTGCTGACGCTGCTCTATTTGGAGGAAACAGGAGCAACCAATGTAT 1677
 QY 491 -----AspAlaSerAlaLeuAspIleArgMetGlyThrLeuLysArgLeuSerProHis 508
 |||||
 Db 1678 GTTGGTCCACCTCACTGACAGACCTCAGCTGCGCAAAATCCAAATGCTCACTCT--- 1734
 QY 509 TyrThrGlnAlaThrSerGluAlaValPheIleHisGluGlyTyr----- 523
 |||||
 Db 1735 -----GAGCATGTGTATTATTCATCCGGAATGGAAGCTGCTGAAGTTC 1776
 QY 524 ---ThrHisAspAlaGlyPheAspAsnAspIleAlaIleLysLeuAsnLysVal 542
 |||||
 Db 1777 CCAGAGAGCAAGCAATTTGTATATGACATTTGCACCTGTGGCGCTGAAGAACCCAGTTC 1836
 QY 543 ValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMet 562
 |||||
 Db 1837 AAATGCGACCCACCGCTCTCCATCTCCATCTGCTGACAGGACCTTTCGCACTCAACCTTC 1896
 QY 563 ArgThrAspAspIleGlyThrAlaSerGlyTyrGlyLeuThrGlnArgGlyPheLeuAla 582
 |||||
 Db 1897 ATGATGGGAGCTGGAGCTGATCTCAGCTGGGCGCAACAGAGAGAGATCTGTCT 1956
 QY 583 ArgAsnLeuMetTyrValAspIleProIleValAspHisGlnLysCys---ThrAlaAla 601
 |||||
 Db 1957 GTTCGCCCTCAAGCGGCAAGCTTACTGTAGCTCTTTAAGAAATGCAAGAGTGA 2016
 QY 602 TyrGluLysProProTyrProArgLysSer-----ValThrAlaAsnMetLeuCysAla 619
 |||||
 Db 2017 GTGGAGAAACCCACAGACAGATGACAGAGCTTATGTTTACCTCTTAACATGATCTGTCT 2076
 QY 620 GlyLeuGluSerGlyLysAspSerCysArgGlyAspSerGlyLysAlaLeuValPhe 639
 |||||
 Db 2077 GGAAGAGAGAG---GGCATGATACCTGTAAAGGGGAGAGTGGTGGGCTTTCCTGTA 2133
 QY 640 LeuAsp---SerGluThrGluArgTrpPheValGlyIleValSerProGlySerMet 658
 |||||
 Db 2134 CAGAGTCCCAATATACAAAGCAAAATTTCTACGACGCTGGCTGTCTCTGGGGCC--- 2190
 QY 659 AsnCysGlyGluAlaGlyGlnTyrGlyValTyrThrLysValIleAsnTyrIleProTrp 678
 |||||
 Db 2191 -----CAGTGTGGACCTATGGGCTCTACACACGGGTAAAGAACTATGTTGACTGG 2241
 QY 679 Ile 679
 |||
 Db 2242 ATA 2244
 |||
 RESULT 9
 US-09-917-800A-1348
 ; Sequence 1348, Application US/09917800A
 ; Patent No. US20020119462A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mendrick, Donna
 ; APPLICANT: Porter, Mark
 ; APPLICANT: Johnson, Kory
 ; APPLICANT: Castle, Arthur
 ; APPLICANT: Elashoff, Michael
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Molecular Toxicology Modeling
 ; FILE REFERENCE: 44921-5038-US
 ; CURRENT APPLICATION NUMBER: US/09/917, 800A
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 607/222, 040
 ; PRIOR FILING DATE: 2000-07-31


```
; Patent No. US2002015515A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Simek, Richard A
; APPLICANT: Heitman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US2002015515A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; OR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 91
; LENGTH: 3318
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-808-602-91

Alignment Scores:
pred. No.: 2,45e-39 Length: 3318
Score: 498.00 Matches: 158
Percent Similarity: 33.67% Conservaive: 73
Best Local Similarity: 23.03% Mismatches: 194
Query Match: 13.25% Indels: 261
DB: 9 Gaps: 20

US-09-874-198-2 (1-686) x US-09-808-602-91 (1-3318)

QY 17 ProLeuGlyProLys-TirProGluProValPheGlyArgLeuAlaSerProGlyPhePr 36
DB 143 CCTCTGGCCCAAGAGACTGCCAG-----CACCTGACATCCCGCGGTACCC 150
QY 36 ocGlyGluTyrAlaAlaAspGlnGluArgArgTrrPrrLeuThAlaProProGlyTyrAr 56
DB 191 AGAGCCGTATGAGCAAGAGCCAGAGACAGACGACATCCAAAGCTCCACAGGGCTTTC 250
QY 56 gLeuArgLeuTyrPheThrHisPheAspLeuGluLeuSerHisLeuGlyGlnGluSerThAs 76
DB 251 TGTGAGGCTGTCTTCCAGAGACTTCCAGCTGAGCCGTCCAGAGACTGTGACAGGGACTC 310
QY 76 eValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuGlyGlnGluSerThAs 96
DB 311 TGTACAAATCTCATTCGTGCGCTTCGATCCAGCAACCAAGTTCTGTGTCAGCAAGCTCCC 370
QY 96 pThrGluAlaGalaProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspIleTh 116
DB 371 TCTGGCAGAGCCCCCTGTGTCAGAGAGGAGTTTGTATCTTCAGGAGAGAGTTTGGCTGAC 430
QY 116 rPheArgSerAspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAlaAl 136
DB 431 CTTCCGACACAGACCTTCTCTCGAG----- 455
QY 136 eGluAspIleAspGluGlnAlaValAlaProGlyGluAlaProThrCysAspHisHscY 156
DB 455 ----- 455
QY 156 sHisAsnHisLeuGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArgAs 176
DB 456 -----AA 457
QY 176 nLysArgThrCysSerAlaLeuGlySerGlyGlnAlaPheThrGlnArgSerGlyGluLe 196
DB 458 CAG----- 461
QY 196 userSerProGluTyrProArgProTyrProLysLeuSerSerCysThrTyrSerIleSe 216
```

```
DB 462 -----ACTGCCCA 469
QY 216 rLeuGluGluGlyPheSerValIleLeuAspPheValGluSerPheAspValGluThrHl 236
DB 470 CCTCCACAAAGGGCTTCGTGCGCTTAC----- 497
QY 236 sProGluThrLeuGlyProTyrTrpPheLeuLysIleGlnThrAspArgGluGluHisGl 256
DB 497 ----- 497
QY 256 yPropheCysGlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThrIi 276
DB 498 -----CAAACTGGCTGT 511
QY 276 eThrPheValThrAspGluSerGlyAspHisThrGlyTrrLysIleHisThrSerTh 296
DB 512 GAACATAT----- 518
QY 296 rAlaGlnProCysProTyrProMetAlaProProAsnGlnHisValSerProValGlnAl 316
DB 519 -AGTCAGCCCATCAGC----- 533
QY 316 aLysTrrIleLeuLysAspSerPheSerIlePheCysGluThrGlyTrrGluLeuLeuGl 336
DB 534 -----GAGCCACGACAG 544
QY 336 nGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLysAspGlySerTrpAspAr 356
DB 545 GGGCTCT----- 551
QY 356 gProMetProAlaCysSerIleValAspGlyProProAspAspLeuProSerGlyAr 376
DB 551 ----- 551
QY 376 gValGluTrrIleThrGlyProGlyValThrThrTrrLysAlaValIleGlnTrrSerCy 396
DB 552 ----GAGGCATTCAGACGACCTGTGACAAACCTGCGCAAG-----CTCCAGAACACAG 601
QY 396 sGluGluThrPheTrrThrMetLysValAsnAspGlyLysTrrValCysGluAlaAspGl 416
DB 602 CCAGAGAGCCCTATATAT---CAGGCCGCGCAGCAGGAGGACATCCACTGTGTCAACCCAG 658
QY 416 yPheTrrPrrHisSerSerGlyGlyLysSerLeuProValCysGluGly--- 435
DB 659 GACCTGGAAGACAGACAGAGATGGAGAGGTTCTTCACTGTATGCTGTGTGCGGAGCG 718
QY 436 -----LeuSerAlaArgThrThrGlyArgIleTrrGlyGlnLys 450
DB 719 GCCAGTCACCCCATTCGCCAGAAATCAGACGACC-----CTGGGTTCCTCCAG 766
QY 450 sAlaLysProGlyAspPheProTrrPrrValLeuIleLeuGlyGlyThrThAlaAlaGl 470
DB 767 AGCCAAAGCTGGGCAACTCCCTGCGCAAGCTTCACACAGTATCCAGCGCGGTGGGGCGG 826
QY 470 yAlaLeuLeuTrrAspAsnTrrPrrValLeuThrAlaAlaHisAlaValTrrGluGlnLys 489
DB 827 GGGCTGTGTGGGGAGAGATGATCTCTCACTGTGCGCACACCGCTTACCCCAAGAGACAG 886
QY 490 -----HisAspAlaSerAlaLeuAspIleArgMetGly-----Th 501
DB 887 TGTTCCTTCAGAGAAACACAGAGTGTGAATGTCTTGGGCAACAGCCATGATATGA 946
QY 501 rLeuLysArgLeuSerProHisTrrThrGlnAlaTrrPrrSerGluAlaValPheIleHisGl 521
DB 947 GATGCTGAACACTGGGCAACACCTGTCTC-----CACCGTGTGTTGTGCACCC 994
QY 521 uGlyTrrThrHisAspAla-----GlyPheAspAsnAspIleAlaLeuIleLysLeuAs 539
DB 995 CGACTACCGTCAGAAATGACTCCCATTAATCTTACGGGAGACATCGCCCTCTGAGGCTGA 1054
QY 539 nAsnLysValValIleAsnSerAsnIleThrProIleCysLeuProArgLysGluAlaGl 559
```



```

Db 486 CCTCCACAAAGGCTCTGGCCCTAC----- 513
QY 236 SPROGLuThLeuCyserProtyrAspPheLeuLysIleGIuThrAspArgIuGIuHISGI 256
Db 513 ----- 513
QY 256 yProphecysGIyLysThrLeuProHISArgIleGIuThrLysSerAsnThrValThrII 276
Db 514 -----CAACCGTGGGTGT 527
QY 276 eHrPheValThAspGIuSerGIyAspHISThrGIyTrpLysIleHISThrSerTh 296
Db 528 GAACATP----- 534
QY 296 rAlaGIuProCyserProtyrProMetaIaProProAsnGIyHISValSerProValGIuAl 316
Db 535 -AGTCAAGCCCATGAC----- 549
QY 316 aLysTyrlLeuLysAspSerPheSerIlePheCysGIuThrGIyTrGIuLeuLeuGI 336
Db 550 -----GAGGCCAGACAG 560
QY 336 nGIyHISLeuProLeuLysSerPheThrAlaValCysGIuLysAspGIySerTrpAspAr 356
Db 561 GGGCTCT----- 567
QY 356 qPrometProAlaCyserIleValAspCysGIyProProAspAspLeuProSerGIyAr 376
Db 567 ----- 567
QY 376 gValGIuTyrlLeThrGIyProGIyValThrThrTyrlLysAlaValIleGIuThrSerCy 396
Db 568 ----GAGGCCATCAAGCAGCTGGAGACAACTGCCAAG-----GTCAGAAACACAG 617
QY 396 sGIuGIuThrPheTyrlThrMetLysValAsnAspGIyLysTyrlValCysGIuAlaAspGI 416
Db 618 CCAAGGAGCCCTATTAT---CAGGCCGGGAGAGGGGAGCAGTCACTGTCACCAACCCAGG 674
QY 416 yHrEtpThrSerSerLysGIyGIuLysSerLeuProValCysGIuProAlaCysGIy-- 435
Db 675 GACCTGGAAGACAGACAGAGATGGGGAGGGGTCTTCACTGATGCTGTGGCGAGC 734
QY 436 -----LeuSerAlaArgThrThrGIyArgIleTyrlGIyLysGIuLys 450
Db 735 GCCAGTCAACCCCATGGCCAGATGACAGACAC-----CTGGTCTTCCAG 782
QY 450 sAlaLysProGIyAspPheProTrpGIuValLeuIleLeuGIyGIyThrThrAlaAlaGI 470
Db 783 AGCCAGAGCTGGGCACTCCCTGGGCAAGCCTTCAACAGATATCCAGGGCGTGGGGCGG 842
QY 470 yAlaLeuLeuTyrlAspAsnTrpValLeuThrAlaAlaHISAlaValTyrlGIuGIuLys-- 489
Db 843 GGGCTGTGCTGGGGACAGATGATCTCTACTGCTGCCACACCATCTACCCCAAGAGACAG 902
QY 490 -----HISAspAlaSerAlaLeuAspIleArgMetGIy-----Th 501
Db 903 TGTTCCTCTCAGAGAACACAGAGTGTGAATGTGTCTTGGGCCACACAGCATATGATGA 962
QY 501 rLeuLysArgLeuSerProHISTyrlThrGIuAlaTrpSerGIuAlaValaPheIleHISGI 521
Db 963 GATGCTGAACACTGGGACACCCCTGTC-----CACCGTGTCTGTGTGCACCC 1010
QY 521 uGIyTyrlThrHISAspAla-----GIyPheAspAsnAspIleAlaLeuLysLeuAs 539
Db 1011 CAGCTACCGTCAGAAATGATGCCATTAATTAGCGGGGAGATGCGCTCTGGAGCTGCA 1070
QY 539 nAsnLysValaValIleAsnSerAsnIleThrProIleCysLeuProArgLysGIuAlaGI 559
Db 1071 GCACACATGCCCTGGGCCCAACAGTCCCGGTGTGTGTGCTGCC-----GATATATGA 1124
QY 559 uSerPheMetArgThrAspAspIleGIyThrAlaSerGIyTrpGIyLeuThrGIuArgGI 579

```

```

Db 1125 GACCTCTTACCGGAGGCGCTGTGGGCTACGTCACTGGGGTTGGCATG--GAGATGGG 1181
QY 579 yPheLeuAlaArgAsnLeuMetTyrlValAspIleProIleValAspHISGIuLysCysTh 599
Db 1182 CTGGCTTACTACTGAGACTGATGACTGAGAGCTGCTGTGATCCCAAGGAGAGCTGCA 1241
QY 599 rAlaAlaTyrlGIuLysProProTyrlProArgLysValThrAlaAsnMetLeuCysAl 619
Db 1242 CGCCTGGCTTCCAAAGAGACAGACAGACCCGAG--GTGTTTCTGACATATATGTGTGT 1298
QY 619 aGIyLeuGIuSerGIyGIyLysAspSerCysArgGIyAspSerGIyGIyAlaLeuValPh 639
Db 1299 TGGGATGAGACCCAAAGGACAGCTGTCTGCCAGGGGAGCAAGTGCAGGCTTATGTGT 1358
QY 639 eLeuAspSerGIuThrGIuArgTrpPheValGIyGIyIleValSerTrpGIySerMetAs 659
Db 1359 ATGGGACATATCATGCCCATATCATCTGGGTGGCCAGCGGCATTTGTCTGGGGC---ATAG 1415
QY 659 nCysGIyGIuAlaGIyGIuThrGIyLysValThrThrLysValIleAsnTyrlleProTrpII 679
Db 1416 GTGTGGCGAAGG-----TATGACTTCTACACCAAGTGTCTCAGCTATGTGACTGTGAT 1469
QY 679 eGIuAsnIleIleSer 684
Db 1470 CAAGGAGTGATGAT 1485

RESULT 15
US-09-808-602-18/c
; Sequence 18, Application US/09808602
; Patent No. US2002015115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougal, John
; TITLE OF INVENTION: NO. US2002015115A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-808-602-18

Alignment Scores:
Pred. No.: 1,43e-39 Length: 1691
Score: 496.00 Matches: 157
Percent Similarity: 33.67% Conservative: 74
Best Local Similarity: 22.89% Mismatches: 194
Query Match: 13.20% Indels: 261
DB: 9 Gaps: 20

US-09-874-198-2 (1-686) x US-09-808-602-18 (1-1691)
QY 17 ProLeuGIyProLys-TrpProGIuProValPheGIyArgLeuAlaSer-ProGIyPhePr 36
Db 1509 CCTCTTGGCCCAAGACTACCCAG-----CAGTGAACATCCCGGGGTACCC 1462
QY 36 oGIyGIuTyrlAlaAsnAspGIuArgArgTrpThrLeuThrAlaProProGIyTyrlAr 56
Db 1461 AGAGCCGTATGGCAAGGCGAAGAGAGCAGACGACATCAAGGCTCCAGAGGCGTTTGC 1402

```

```

QY 56 gLeuArgLeuTyrPheThrHisPheAspLeuGluLeuSerHisLeuCysGluTyrAspPh 76
   :::::::::: :::::::::: :::::::::: ::::::::::
Db 1401 TGTGAGGCTGTCTTCAGAGACTTGACCTGACCTCCAGAGACTGTCAGAGGACTC 1342
QY 76 eValLysLeuSerSerGlyAlaLysValLeuAlaThrLeuCysGlyGlnLeuSerThra 96
   :::::::::: :::::::::: :::::::::: ::::::::::
Db 1341 TGTCAAACTCTCATCTGTGCTGCGATCCAGACAGCTTCTGTGTCAGACAGCTCC 1282
QY 96 pThrGluArgAlaProGlyLysAspThrPheTyrSerLeuGlySerSerLeuAspLeuTh 116
   :::::::::: :::::::::: :::::::::: ::::::::::
Db 1281 TCTGGGCGAGCCCCCTGGTCAAGAGGAGATTGTATCTCAGAGAGATTGGCGCTGAC 1222
QY 116 rPheArgSerAspTyrSerAsnGluLysProPheThrGlyPheGluAlaPheTyrAla 136
   :::::::::: :::::::::: ::::::::::
Db 1221 CTTCGCCACACAGCTTCTCCGAG----- 1197
QY 136 aGluAspLLeuAspGluCysGlnValAlaProGlyGluAlaProThrCysAspHisHscY 156
   :::::::::: :::::::::: ::::::::::
Db 1197 ----- 1197
QY 156 sHisAsnHisLeuGlyGlyPheTyrCysSerCysArgAlaGlyTyrValLeuHisArg 176
   :::::::::: :::::::::: ::::::::::
Db 1196 -----AA 1195
QY 176 nLysArgThrCysSerAlaLeuCysSerGlyGlnValPheThrGlnArgSerGlyGlu 196
   :::::::::: ::::::::::
Db 1194 CAAG----- 1191
QY 196 uSerSerProGluTyrProArgProTyrProLysLeuSerSerCysThrTyrSerIle 216
   :::::::::: ::::::::::
Db 1190 -----ACTGCCCA 1183
QY 216 rLeuGluGluGlyPheSerValLLeuLeuAspPheValGluSerPheAspValLunThr 236
   :::::::::: ::::::::::
Db 1182 CCTCCAAAGGCTTCTGCGCTCTAC----- 1155
QY 236 sProGluThrLeuCysProTyrAspPheLeuLysIleGlnThrAspArgGluLunHisG 256
   :::::::::: ::::::::::
Db 1155 ----- 1155
QY 256 yProPheCysGlyLysThrLeuProHisArgIleGluThrLysSerAsnThrValThr 276
   :::::::::: ::::::::::
Db 1154 -----CAAACCTGCGCTGT 1141
QY 276 eThrPheValThrAspGluSerGlyAspHisThrGlyTyrLysIleHisTyrThrSer 296
   :::::::::: ::::::::::
Db 1140 GAACATA----- 1134
QY 296 rAlaGlnProCysProTyrProMetaLarProAsnGlnHisValSerProValGln 316
   :::::::::: ::::::::::
Db 1133 -ACTCAGCCCATCAGC----- 1119
QY 316 aLysTyrLLeuLeuLysAspSerPheSerIlePheCysGluThrGlyTyrGlnLeuGlu 336
   :::::::::: ::::::::::
Db 1118 -----GAGGCCAGCAG 1108
QY 336 nGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLysAspLysSerTrpAsp 356
   ::::::::::
Db 1107 GGGCTCT----- 1101
QY 356 gProMetProAlaCysSerIleValAspCysGlyProProAspAspLeuProSerGly 376
   ::::::::::
Db 1101 ----- 1101
QY 376 gValGluTyrLLeuThrGlyProGlyValThrTyrLysAlaValIleGlnTyrSer 396
   :::::::::: ::::::::::
Db 1100 ---GAGGCACTCAAGCACTGAGACAACCTGCCAG-----GTCCAGAACACTG 1051
QY 396 sGluGluThrPheThrThMetLysValAsnAspGlyLysTyrValCysGlnLunLasp 416
   :::::::::: ::::::::::
Db 1050 CCAAGAGCCCTATTAT---CAGGCCCGGAGAGGAGGAGCACTGACCTGCAACCCAG 994
QY 416 yPheTyrThrSerSerLysGlyLunLysSerLeuProValCysGluProValCysGly 435

```

```

Db 993 GACCTGAAAGACAGACAGAGATGGGAGAGTTCCTCACTGATGAGCTGTGCGGAG 924
QY 436 -----LeuSerAlaArgThrThrGlyGlyArgIleTyrGlyGlnLun 450
   :::::::::: ::::::::::
Db 933 GCCAGTCACCCCATTCGCCAGATGACAGCAGC-----CTCGTTCTTCCAG 886
QY 450 sAlaLysProGlyAspPheProThrGlnValLeuIleLeuGlyGlyThrThrAlaLag 470
   :::::::::: ::::::::::
Db 885 AGCCAAGCTGGGCACTCCCTGGCAAGCTTCACCAAGTATCCAGGCGGCGG 826
QY 470 yAlaLeuLeuTyrAspAsnThrValLeuThrAlaAlaHisAlaValTyrGlnLun 489
   :::::::::: ::::::::::
Db 825 GGGCTGTCTGGGAGACAGATGATCTCACTGCTGCCACACAGCATGACCCAGAGCAG 786
QY 490 -----HisAspAlaSerAlaLeuAspLLeuArgMetGly-----Th 501
   :::::::::: ::::::::::
Db 765 TGTTCCTTCAGAGAACCCAGATGTGAATGTCTTCTGGCCACACAGCCATGATGA 706
QY 501 rLeuLysArgLeuSerProHisTyrThrGlnAlaTyrPserGluAlaValPheIleHis 521
   :::::::::: ::::::::::
Db 705 GATGCTGAACCTGGGAGACCACTCTGC-----CACCGTGTCTTGTGCACCC 658
QY 521 uGlyTyrThrHisAspAla-----GlyPheAspAsnAspLLeuAlaLeuLysLeu 539
   :::::::::: ::::::::::
Db 657 CGACTACCCGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 598
QY 539 nAsnLysValValLLeuAsnSerAsnIleThrProLysCysLeuProArgLysGlu 559
   :::::::::: ::::::::::
Db 597 GCACAGATCTCCCTGGGCCCCCAACGTCCTCCGCTGTCTGCC-----GATATGA 544
QY 559 uSerPheMetArgThrAspAspLLeuGlyThrAlaSerGlyTyrGlyLeuThrGlnArg 579
   :::::::::: ::::::::::
Db 543 GACCTCTACCGCAGCGGCTTGTGGCTACGTCAGAGGAGGTTGGCAGT---GAGATGG 487
QY 579 yPheLeuAlaArgAsnLeuMetTyrValAspLLeuProIleValAlaAspHisGlnLun 599
   :::::::::: ::::::::::
Db 486 CTGGCTAACTAGTACGTGATGATGATGATGATGATGATGATGATGATGATGATG 427
QY 599 rAlaAlaTyrGlnLysProProTyrProArgGlySerValThrAlaAsnMetLeu 619
   :::::::::: ::::::::::
Db 426 CGCCTGCTCCAAAGAGACAGAGACCCGAG---GTGTTTCTGACAAATGTCTCTGT 370
QY 619 aGlyLeuGluSerGlyLysAspSerCysArgGlyAspSerGlyGlyLunLunValPh 639
   :::::::::: ::::::::::
Db 369 TGGGATGAGACCAAGAGCAGATGCTGCCAGGGGAGAGTGGCAGCTTATGTGT 310
QY 639 eLeuAspSerGluThrGluArgTyrPheValGlyGlyIleValSerTrpLysSerMet 659
   :::::::::: ::::::::::
Db 309 ATGGACAAATCATGCCATCATCGGTGGCCAGGGGATTTGTCTCGGGGC---ATAGG 253
QY 659 nCysGlyGluAlaGlyGlnTyrValTyrThrLysValLLeuAsnTyrLLeuProThr 679
   :::::::::: ::::::::::
Db 252 GTCTGGGAGAGG-----TATGACTTCTACACCAAGGTCCTACGATATGTGACTGAT 199
QY 679 eGluAsnLLeuLeuSer 684
   ::::::::::
Db 198 CAAGGAGATGATGAAT 183

```

RESULT 16
 US-10-099-700A-1
 : Sequence 1, Application US/10099700A
 : Publication No. US20030008372A1
 : GENERAL INFORMATION:
 : APPLICANT: Edwin L. Madison
 : APPLICANT: Edgar O. Ong
 : TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE
 : FILE REFERENCE: 24745-1613
 : CURRENT APPLICATION NUMBER: US/10/099,700A
 : PRIOR FILING DATE: 2002-05-24
 : PRIOR APPLICATION NUMBER: 60/275,592
 : PRIOR FILING DATE: 2001-03-13

Oy	9	LeuLeucYsgSLysSerValAlaThrProLeuGlyProLysTyrProGlu	-----	24
Db	636	GTGCTAACATCTGTGTGGCTTCCCATTTGACCCAGAAATGTCGACAGAGACTCAGAC	695	
Oy	25	-----ProValPheGlyArgLeuAlaSerPro	33	
Db	696	AACAGCTCAGCTTTTGCCCTGCATGCCCATGGTGCAGCAGTACACGGTTACTACACCT	755	
Oy	34	GlyPheProGlyGlu--TyrAlaAsnAspGlnAlaArgTyrPheThrLeuThrAlaPro	52	
Db	756	GGCTTCCCAACAGTCCCTACCCGGGGCATGCCCCGCTGCAGTGGTCTCGGGGGGAC	815	
Oy	53	ProGlyTyrArgLeuArgLeuTyrPheThrHisPheAspLeuGlnLeuSerHisLeucys	72	
Db	816	GCCGACTCTGTGCGACCTCCACCTCCGAAAGTGTGATGTGCTGCTCGTGTGATGAGAT	875	
Oy	73	GluTyrAspPheValLysLeuSerSerGlyAlaLys-----ValLeuAla	87	
Db	876	GGCAGTGCCTGGTACCCGGTATGATAGCCTGAGCCCCCATGAAACCCACGCTGGTG	935	
Oy	88	ThrLeuLysGlnGlnSerThrAspThrGluArgAlaProGlyLysAspPheTyr	107	
Db	936	CGGCTGTGTGCACCTTCTCACCTCTCAACCTG-----ACTTCTCTC	980	
Oy	108	SerLeuGlySerSerLeuAspLleThrPheArgSerAspTyrSerAsnGlySerProPhe	127	
Db	981	TCTCTCCAGAACGTCTTCTTGACGCGTGATACAAATCTGACGCGGCGACATCT--	1037	
Oy	128	ThrGlyPheGluAla--PheTyrAlaAlaGluAspLleAspGlyCysGlnValAlaPro	146	
Db	1038	---GGCTTTAGGCGCACTTCTTCCAGCTGCCCAAGATAGACGTG-----	1082	
Oy	147	GlyLeuAlaProThrCysAspHisHisCysHisAsnHisLeuGlyLysPheTyrCysSer	166	
Db	1083	-----GGCGGCTT-----	1091	
Oy	167	CysArgAlaGlyTyrValLeuHisArgAsnLysArgThrCysSerAlaLeuCysSerCyl	186	
Db	1091	-----	1091	
Oy	187	GlnAlaPheThrGlnArgSerGlyLysLeuSerSerProGlyTyrProArgProTyrPro	206	
Db	1092	-----TTGAGTGCACCCCAAGGAGACATTTAGCAGCCCTTACTATCCAGGCCACTACCG	1145	
Oy	207	LysLeuSerSerCysThrTyrSerLleSerLeuGlnGlyLysPheSerValLleLeuAsp	226	
Db	1146	CCCAACATCACTGTCACATGGATATCAAGTGTCCCAACACGGAGACGTGAAGTGGCC	1205	
Oy	227	PheValGluSerPheAspValGluThrHis--ProGluThrLeuCysProTyrAspPhe	245	
Db	206	TTTCAAACTCTTCTTGTGTGGACCCCAAGTACAGTGGTCTCTGACCAAGAGACTAT	1265	
Oy	246	LeuLysLleGlnThrAspArgGluGlnHisGlyProPheCysGlyLysThrLeuProHis	265	
Db	1266	GTTGGAGATCAACGGGGGAGAG-----TACTGGCGTGAAGAGTCCACAGTTT	1310	
Oy	266	ArgLleGluThrLysSerAsnThrValThrLleThrPheValThrAspLysSerGlyAsp	285	
Db	1311	GTGGTGAAGCAGCAACAGCAGACAGATTAACAGTCCACTTCCATTGTGATCATCGTCACAG	1370	
Oy	286	HisThrGlyTyrPylLysLleHisTyrThrSer-----ThrAlaGlnProCysProTyrPro	303	
Db	1371	GACACCGGGTCTCAGCTGAGTACCTCTCCAGTCCACAGACACCGGCCCCAGGGATG	1430	
Oy	304	MetAlaProProAsnGlyHisValSerProValGlnAlaLys-----	317	
Db	1431	TTTCATGTGCAGACTGGAGCGTGCATCCGAAGAACTCGCTGCGAGCGCTGGGACAGC	1490	
Oy	318	-----TyrLleLeuLysAspSerPheSerLlePheCysGlyUthrGlyTyrGlnLeu	334	
Db	1491	TGCCCGGATTAT-----AGTGATGAGCGTTACTGCCGATGCAATGCCACCCAGCATGTC	1544	
Oy	335	LeuGlnGlyHisLeuProLeuLysSerPheThrAlaValCysGlnLys-----	350	

Db	1545	ACGTGCAAAAACCAAGTTCATGCAAGCCCTCTTGTGGTGTGTGACAGTGTCAACAGACTGT	1604
Oy	351	---AspGlySerTrpPaspArg-----PrometProAla-----CysSerIle	363
Db	1605	GGGGACGGAAGTGCACGAGGAGGGCTGCAGCTGCTCTGGGAGTTTCAAGTGTCCAAAT	1664
Oy	364	ValAspCysGlyProProPaspAspIleuProSerGlyArgValGluTyrIleIlethGlyPro	383
Db	1665	GGGAAGCTCTCCCTCGACGACCAGAACTGAATGGGAAGACACTGTGGAGATGGGCT	1724
Oy	384	GlyValIThrTrpTrpLysAlaValIleGlnTyrSerCysGluGluThrPheTyrThrMet	403
Db	1725	GACACAGCTTCATGTGTGACACCTGGAATGTCTCTTCC-----ACCAATATATCTTAC	1778
Oy	404	LysValAsnAspGlyLysTyrValCys-----GluAlaAspGlyPhe	417
Db	1779	CGCTGCCCAAAATGGC-----CTCTGTGTGAGCAACGGCAACCTCGAGTGTGATGGGAAG	1832
Oy	418	TrpThrSerSerLysGly-----GluLysSerLeuProValCysGluProValCysGly	435
Db	1833	ACGGAAGCTGATCGGCTCGCATGGAGAAAAC-----TGTGAC-----TGTGGG	1877
Oy	436	LeuSerAlaArgTrpThrProGlyArgGlyIleTyrGlyGlyAlaLysProGlyAsp	455
Db	1878	CTGGCATCTCTTTACCAACAGGCTCGGTGGTTGGTGTGGACCAATGGCAGAGGGCCAG	1937
Oy	456	PheProTrpGlnValLeuIle-----LeuGly---GlyThrThAlaAlaGlyAlaLeu	472
Db	1938	TGGCCCTGGCAGTGCAGCCCTCCACAGCCCTGGCCAGGCGCACTGTGTGGGGCTCGCTC	1997
Oy	473	LeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyrGluGlnLysHis-----	490
Db	1998	ATCTCTCTGACTGGCTGGTCTCTGTGCAGGCTCATTTGCTTCAGAGTACCAAAAATTTCAG	2057
Oy	491	-----AspAlaSerAlaLeuAspIleArgMetGlyThrLeu-----LysArgLeu	505
Db	2058	TACTCAGACTACACAGATGTGACGGCTTCTGGGTCTGTCTGTGACCAGCAAGGGCAGT	2117
Oy	506	SerProHisTyrTrpGlnAlaTrpSerGlyAlaValPheIleHisGluGlyTyrThrHis	525
Db	2118	GCCCTGGGGGCGAGGAGCTGCAAGCTCAACAGTATCATACCACCCCTTCCTC---AAT	2174
Oy	526	AspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValValIleAsn	545
Db	2175	GATTTACACTTCGCATGTGATCATCGCTTGTGAGACTGGAAGATCGGTGGAGTACAGC	2234
Oy	546	SerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThrAsp	565
Db	2235	ACCGCTGGCCGCCCATCTCCCTGCTCCTC-----GATGCAACCATATGCTTCCCTGCTGCC	2288
Oy	566	AspIleGlyThrAlaSerGlyTyrTrpGlyLeuThrGlnArgGlyPheLeu---AlaArgAsn	584
Db	2289	AAGGCCATCTGGGTCACAGGCTGGGGGCAACAAAAAGAGGAGTATCCGAGCCGTATC	2348
Oy	585	LeuMetTyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrGluLys	604
Db	2349	CTGCAGAAAGGGGTAGATCCGTCTCATTCACCAACACACCATCTGAGACCTCATGCTCCG	2408
Oy	605	ProProTyrProArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerCly	624
Db	2409	-----CAGATTCACCCCAACGATATGTGTGGTTCCTCATGTGG	2450
Oy	625	GlyLysAspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerLThr	644
Db	2451	GGTGTGCACTCTCTCCACAGGGTGACTGTGGTGGCCCTTGTCAAGCGCGAGAAACAT---	2507
Oy	645	GluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsnCysGlyGluAlaGly	664
Db	2508	GGGGCAATGTTCCAGGCTGGTGTGGTGTAGCTGGGGTGA---GCGTGGCCCTCAGAGAAC	2564
Oy	665	GlnTyrGlyValTyrThrLysValIleAsnTyrIleProTrpIleGlu	680


```

Db 2565 AAGCAGGCGTGTACACAGAGCTCCCTGTATTGGGAGACTGGATCAAA 2612

RESULT 20
US-09-981-123-1
: Sequence 1, Application US/09981123
: Patent No. US20020151029A1
: GENERAL INFORMATION:
: APPLICANT: Holloway, James L.
: TITLE OF INVENTION: Human Serine Protease
: FILE REFERENCE: 99-88C1
: CURRENT APPLICATION NUMBER: US/09/981,123
: CURRENT FILING DATE: 2001-10-16
: PRIOR APPLICATION NUMBER: 60/167,038
: PRIOR FILING DATE: 1999-11-23
: PRIOR APPLICATION NUMBER: 09/715,994
: PRIORITY DATE: 2000-11-17
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FASTSEQ for Windows Version 3.0
: ID NO 1
: LENGTH: 807
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(807)
US-09-981-123-1

Alignment Scores:
Pred. NO.: 3,72e-35 Length: 807
Score: 446.50 Matches: 100
Percent Similarity: 53.01% Conservative: 41
Best Local Similarity: 37.59% Mismatches: 92
Query Match: 11.88% Indels: 33
DB: 10 Gaps: 8

US-09-874-198-2 (1-686) x US-09-981-123-1 (1-807)
OY 444 ArgIeYrGyIgYgLnLsAlaLysPrGgIyAspPhrPrGrrgInValLeuIleu 463
||||| ||||||| :|||:||||| ||||||| |||||||
Db 1 AGCATCATGGTGGGGCCGAAGATGCTGAGCTGCTTCCCGTGGCAGGCCCTGATG 60
OY 464 GlyeIYrThrAla-----AlaGlyAlaLeuIleYr 474
||||| :|||:|||||
Db 61 GTGGAGCAGCTTGCAGAGTCCCAATGACAAAGTGTTGGAGATGGGGCCCTGCT 120
OY 475 AspaNrPrVAlLeuThrAlaAlaHIsAlaValYrGluGlnYrSHsAspAlaSer 494
|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 GCGTCCTGGATCTTCACAGCAGCTCATGTGGCGCTCCAGCGCTAGACACCGCGTG 180
495 LeuAspIle--ArgmeIgIYrThLeuLysArgLeuSerPrOHsIYrThrGlnAla 513
::: :::: ||| ||| ::::
Db 181 ATACAGCTCTCCAGAGCATGTCACCGCTCTACTGGCGTTCGTATGTCGCGAG 210
::: :::: ||| ||| ::::
OY 514 SerGluAla-----ValPheIleHsIeGluGlyYrThrHisAsp 526
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TCGGGGCGAGTCAACAGCTCAGCTGCCCGAGATGGTGGCTCCACCCAGACTTC--AACATC 297
OY 527 AlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnYsValValIleAsnSer 546
298 CAAAACTTACCAACCCAGATATGCTCTGTGGTGAAGCTGAGGAGCGCTGGCCCTGGAGCC 357
547 AsnIleThrPrIleCysLeuPrArgLysGluIaGluLysPheMetArgThrAspAsp 566
358 CACGTTATGCTGTCTGCTGCTGCCAAGGCTTAGGCTTAAGC---CCGGCCCCCACATG 614
OY 567 IlegIYrAlaSerGlyTTPeIYrLeuThrGlnArgIYrPhe-----580
:::||||| ::|||:|||||:|||||:|||||:|||||:|||||:|||||
Db 415 CTGGGCGTGTGGCGGCTGGGCGATCTCCAAATCCCATGTCGACAGTGATGATCATC 474
581 -----LeuAlaArgAsnLeuMetYrValAspIlePrIlePrIleValAsp 594
475 AGCACTGGCACAGGACCTTGTCAATGTCTCTGCAGTATGTCAAGTAACTTACCGGTGGTCTT 534

```

```

Oy 595 HisGlnIysCysThrAlaAlaTyrGlnIlysProPrtOtyrProAlaGlySerValThrAla 614
Db 535 CACCCTAGTGTCAAAACTACTACTAGATGCCCTGC---GGCAATTACAGCGCTCAGCGAG 591
Oy 615 AsnMetLeuCysAlaGlyLeuGlnSerGlyIlysAspSerCysArgIlysAspSercly 634
Db 592 AACATGTTCTGTGCTGGCTACTACAGAGGGCGGAAGACACAGCGCTTGGATAGTCGGT 651
Oy 635 GlyAlaLeuValPheLeuAspSerpSerGlyThrGluArgTPrpPheValIcylGlyIleValSer 654
Db 652 GGGCGCTTTGTCATCTTTGATGACTTGAGCCACCGCTGGCGTGTCGACAGGCTGGTGTC 711
Oy 655 TrpGlySer---MetAsnCysGlyGluAlaGlyIcylGlyGlyValTyrThrIlysValIle 673
Db 712 TGGGGGGGAGCTGAAGAATGCGCGACAGCACAGCTGTATGTAGACTTACACAAAGTCTCC 771
Oy 674 AsnTyrIlePrtPrtIle 679
Db 772 AATTACGTGAGCTGGGTG 789

RESULT 21
US-09-888-615-54
: Sequence 54, Application US/09888615
: Patent No. US20020064856a1
: GENERAL INFORMATION:
: APPLICANT: PLOMMAN, GREGORY
: APPLICANT: WHYTE, DAVID
: APPLICANT: CAENEPEEL, SEAN
: APPLICANT: CHARYDCZAK, GLEN
: APPLICANT: MANNING, GERARD
: APPLICANT: SODARSANAM, SUCHA
: TITLE OF INVENTION: NOVEL PROTEASES
: FILE REFERENCE: 038602/1214
: CURRENT APPLICATION NUMBER: US/09/888,615
: CURRENT FILING DATE: 2001-06-26
: PRIOR APPLICATION NUMBER: 60/214,047
: PRIOR FILING DATE: 2000-06-26
: NUMBER OF SEQ ID NOS: 150
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 54
: LENGTH: 2409
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-888-615-54

Alignment Scores:
Pred. No.: 3 31e-34 Length: 2409
Percent Similarity: 444.00 Matches: 153
Percent Similarity: 45.00% Conservative: 81
Best Local Similarity: 29.42% Mismatches: 200
Query Match: 11.81% Indels: 86
DB: 10 Gaps: 31

US-09-874-198-2 (1-686) x US-09-888-615-54 (1-2409)
Oy 194 GlyLeuLeuSerSerProGluTyrProArgProTyrProIlysLeuSerSerCysThrTyr 213
Db 1018 GGGCGCTGCACACACCGGTACTTCCACGTACTACCGGCCAACACCACTGCTCTGG 1077
Oy 214 SerIle-----SerLeuGlnGluGlyPheSerValIleLeuAspPheVal----- 228
Db 1078 CACGTCAAGGTGCGCTCTGTGACTACGCGCTTGCGCTGTGATGATGATGATGATGATG 1133
Oy 229 -----GluSerPheAspPheArgIleThrHisProGluThrLeuCysProTyrIleAspPheLeu 246
Db 1138 AGAGAGCGGAAGTATGATTTGGCG-----TGCACCCAGGCGCCACTG 1177
Oy 247 LysIleGlnThrAspArgGluGlnHisGlyProPheCysGly-----LysThr 262
Db 1180 ACGATTCGACAGCAGAGG-----CTGTGTGGCTTGGCGCATCTCGCAGGCC 1224
Oy 263 LeuProHisArgIleGluThrLysSerAsnThr---ValThrIleThrPheValThrAsp 281

```

Db 1225 TACGCCGAGAGATCCCGTGTGGCCACGCGCGGATCACAATTCACTCCAG 1284
QY 282 GUSerGlyAspHisThrGlyTrpLysIleHisTyr-----ThrSerThrAlaGlnPro 299
Db 1285 ATCTCCCTCACCGGGCCGGGTGGCGGTGCTACTATGCTTGTACACAGCTGGACCCC 1344
QY 300 CysProTyrPro---MetalapProAsnGlyHisValSerProValAlaGlnLysTyr 318
Db 1345 TGGCCGTGAGAGTTCCTCTTCTGTGAATGAGACTGTCCTCCCT-----GGCTGTGAT 1398
QY 319 IleuLysAspSerPheSerIlePheCysGluThrGlyTrpGluLeuGlnGlyHis 338
Db 1399 GGGGTACAGGAC-----TGCCTCCACAGGC----- 1422
QY 339 LeuProLeuLysSerPheThrAlaValCysGln-----LysAspGlySer 353
Db 1423 -----CTGGATGAGAGAACTGCTTTGAGAGCCACATTCCATGCAAGAGCAGC 1476
QY 354 TrpAspArgProMetProAlaCysSerIleValAspCysGlyProProAspAlaLeuPro 373
Db 1477 ACATGCATCTCACTGCC-----AAGGTCTGAT---GGCAGCCTGATTTGTCTC--- 1524
QY 374 SerGlyArgValAluTrpIleThrGlyProGlyValThrThrTyrLysAlaValIleGln 393
Db 1525 AACGCGAGCGATGAGAGCAGTGCACGAGAGGGGTGCCATGT-----GGACATTTCACC 1578
QY 394 TyrSerCysGluGluThrPheTyrThrMetLysValAsn-----AspGlyLysTyr 410
Db 1579 TTCACGTGTAGAGACGGAGCTGCTGAGAGACCCACCCGCACTGTGTAGGGCGGCC 1638
QY 411 ValCysGluAlaAspGlyPheTrpThrSerSerLysGlyLysSerLeuProValCys 430
Db 1639 GACTGCAGG---GACGCG-----TCGGATGAGAGAC-----TGT 1671
QY 431 GluProValCysGlyLeuSerAlaArgThrThrGlyLysArgIleTyrGlyGlnLys 450
Db 1672 GAC-----GTGGCTCCAGAGGCCCTCCAGC-----CGCATTTGTGTGGAGCTGTG 1719
QY 451 AlaLysProGlyAspPheProTrpGln-----ValLeuIleLeuGlyGlyThrThrAla 468
Db 1720 TCCTCCGAGGTGATGGCTGATGGCAGGCGCAGCTCAGGTTGGGGTCCACACATCTGT 1779
QY 469 AlaGlyAlaLeuLeuTyrAspAsnTrpValLeuThrAlaAlaHisAlaValTyrGln 488
Db 1780 GGGGGGCGCTCATCGCTGACCTGCGGTGATTAACAGCTCCACCTCCAGGAGGAC 1839
QY 489 LysHisAspLysSerAlaLeu---AspIleArgMetGlyThrLeuLysAlaGluSer--- 506
Db 1840 AGCATGGCTCCACGGTCTGTGAGCCGTTCTCTGGCAAGGTGTGCGAGAACTCGCGC 1899
QY 507 ---ProHisTyrThrGlnAlaTrpSerGluAlaValPheIleHisGluGlyTyrThrHis 525
Db 1900 TGGCTGTGAGAGTGTCTTCAAGGTGAGCGCGCTGCTCTGCACCG---TTCACAGAA 1956
QY 526 AspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValIleAsn 545
Db 1957 GAGGACAGCATGACTACAGCTGCGGCTGCTGACCTGCACACCGCGTGGCGCTG 2016
QY 546 SerAsnIleThrProIleCysLeuProArgLysGluAlaLeuSerPheMetAlaGlyThrAsp 565
Db 2017 GCGGCGGTGGCGCGTCTGCTGCGCGCGC-----TCCCATTTCTTGAAGCGCGC 2070
QY 566 AspIleGlyThrAlaSerGlyTrpGly---LeuThrGlnArgGlyPheLeuAlaArgAsn 584
Db 2071 CTCGACTGCTGGATTTACGGCTGGCGCTTTCGCGAGGGCGGCCCATACACACGCT 2130
QY 585 LeuMetTyrValAspIleProIleValAspHisGlnLysCysThrAlaAlaTyrLys 604
Db 2131 CTCGACAAAGTGTGATGCGATGATCCACAGACCTGTGACGCGAGCGCTAT----- 2184
QY 605 ProProTyrTrpArgGlySerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGly 624
Db 2185 -----CGTTACACAGGTGACGCCACGATGCTGTGTGCGGCTACCCCAAGGCGC 2232

QY 625 GlyAspSerCysArgGlyAspSerGlyGlyAlaLeuValPheLeuAspSerGluThr 644
Db 2233 AAGAAGATGCTCTGATGAGGTGATCAGGTGCTCCCTGTG---TGAAGGACATCACT 2289
QY 645 GluArgTrpPheValGlyGlyIleValSerTrpGlySerMetAsnCysGlyAlaGly 664
Db 2290 GCGCGCTGTCTGCGCGGCGGTGCTGACGTGGGC---CTGGAGTGTGCGCGGCTTAC 2346
QY 665 GlnTyrGlyValTyrThrLysValIleAsnTyrIleProTrpIleGluAsnIleIleSer 684
Db 2347 TACTTGGGCTTACACCCGATCACAAGGTGTGATCAGCTGATCCAGCAAGTGTGACC 2406
RESULT 22
US-09-978-295A-168
Sequence 168, Application US/09978295A
Patent No. US2002015606A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James:
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC11
CURRENT APPLICATION NUMBER: US/09/978,295A
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/074750
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886

;;
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22

;;
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084537
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Alignment Scores:
Pred. No.: 4.9e-34 Length: 3143
Score: 444.00 Matches: 153

1	CURRENT APPLICATION NUMBER: 2001-10-16
2	CURRENT FILING DATE: 2001-10-16
3	PRIOR APPLICATION NUMBER: 09/918585
4	PRIOR FILING DATE: 2001-07-30
5	PRIOR APPLICATION NUMBER: 60/062250
6	PRIOR FILING DATE: 1997-10-17
7	PRIOR APPLICATION NUMBER: 60/064249
8	PRIOR FILING DATE: 1997-11-03
9	PRIOR APPLICATION NUMBER: 60/065311
10	PRIOR FILING DATE: 1997-11-13
11	PRIOR APPLICATION NUMBER: 60/066364
12	PRIOR FILING DATE: 1997-11-21
13	PRIOR APPLICATION NUMBER: 60/077450
14	PRIOR FILING DATE: 1998-03-10
15	PRIOR APPLICATION NUMBER: 60/077632
16	PRIOR FILING DATE: 1998-03-11
17	PRIOR APPLICATION NUMBER: 60/077641
18	PRIOR FILING DATE: 1998-03-11
19	PRIOR APPLICATION NUMBER: 60/077649
20	PRIOR FILING DATE: 1998-03-11
21	PRIOR APPLICATION NUMBER: 60/077791
22	PRIOR FILING DATE: 1998-03-12
23	PRIOR APPLICATION NUMBER: 60/078004
24	PRIOR FILING DATE: 1998-03-13
25	PRIOR APPLICATION NUMBER: 60/078886
26	PRIOR FILING DATE: 1998-03-20
27	PRIOR APPLICATION NUMBER: 60/078936
28	PRIOR FILING DATE: 1998-03-20
29	PRIOR APPLICATION NUMBER: 60/078910
30	PRIOR FILING DATE: 1998-03-20
31	PRIOR APPLICATION NUMBER: 60/078939
32	PRIOR FILING DATE: 1998-03-20
33	PRIOR APPLICATION NUMBER: 60/079294
34	PRIOR FILING DATE: 1998-03-25
35	PRIOR APPLICATION NUMBER: 60/079656
36	PRIOR FILING DATE: 1998-03-26
37	PRIOR APPLICATION NUMBER: 60/079664
38	PRIOR FILING DATE: 1998-03-27
39	PRIOR APPLICATION NUMBER: 60/079689
40	PRIOR FILING DATE: 1998-03-27
41	PRIOR APPLICATION NUMBER: 60/079663
42	PRIOR FILING DATE: 1998-03-27
43	PRIOR APPLICATION NUMBER: 60/079728
44	PRIOR FILING DATE: 1998-03-27
45	PRIOR APPLICATION NUMBER: 60/079786
46	PRIOR FILING DATE: 1998-03-27
47	PRIOR APPLICATION NUMBER: 60/079920
48	PRIOR FILING DATE: 1998-03-30
49	PRIOR APPLICATION NUMBER: 60/079923
50	PRIOR FILING DATE: 1998-03-30
51	PRIOR APPLICATION NUMBER: 60/080105
52	PRIOR FILING DATE: 1998-03-31
53	PRIOR APPLICATION NUMBER: 60/080107
54	PRIOR FILING DATE: 1998-03-31
55	PRIOR APPLICATION NUMBER: 60/080165
56	PRIOR FILING DATE: 1998-03-31
57	PRIOR APPLICATION NUMBER: 60/080194
58	PRIOR FILING DATE: 1998-03-31
59	PRIOR APPLICATION NUMBER: 60/080327
60	PRIOR FILING DATE: 1998-04-01
61	PRIOR APPLICATION NUMBER: 60/080328
62	PRIOR FILING DATE: 1998-04-01
63	PRIOR APPLICATION NUMBER: 60/080333
64	PRIOR FILING DATE: 1998-04-01
65	PRIOR APPLICATION NUMBER: 60/080334
66	PRIOR FILING DATE: 1998-04-01
67	PRIOR APPLICATION NUMBER: 60/080370
68	PRIOR FILING DATE: 1998-04-08
69	PRIOR APPLICATION NUMBER: 60/081049
70	PRIOR FILING DATE: 1998-04-08
71	PRIOR APPLICATION NUMBER: 60/081071
72	PRIOR FILING DATE: 1998-04-08
73	PRIOR APPLICATION NUMBER: 60/081195

1	PRIOR FILING DATE: 1998-04-08
2	PRIOR APPLICATION NUMBER: 60/081203
3	PRIOR FILING DATE: 1998-04-09
4	PRIOR APPLICATION NUMBER: 60/081229
5	PRIOR FILING DATE: 1998-04-09
6	PRIOR APPLICATION NUMBER: 60/081955
7	PRIOR FILING DATE: 1998-04-15
8	PRIOR APPLICATION NUMBER: 60/081817
9	PRIOR FILING DATE: 1998-04-15
10	PRIOR APPLICATION NUMBER: 60/081819
11	PRIOR FILING DATE: 1998-04-15
12	PRIOR APPLICATION NUMBER: 60/081952
13	PRIOR FILING DATE: 1998-04-15
14	PRIOR APPLICATION NUMBER: 60/081838
15	PRIOR FILING DATE: 1998-04-15
16	PRIOR APPLICATION NUMBER: 60/082568
17	PRIOR FILING DATE: 1998-04-21
18	PRIOR APPLICATION NUMBER: 60/082565
19	PRIOR FILING DATE: 1998-04-21
20	PRIOR APPLICATION NUMBER: 60/082704
21	PRIOR FILING DATE: 1998-04-22
22	PRIOR APPLICATION NUMBER: 60/082804
23	PRIOR FILING DATE: 1998-04-22
24	PRIOR APPLICATION NUMBER: 60/082700
25	PRIOR FILING DATE: 1998-04-22
26	PRIOR APPLICATION NUMBER: 60/082797
27	PRIOR FILING DATE: 1998-04-22
28	PRIOR APPLICATION NUMBER: 60/082796
29	PRIOR FILING DATE: 1998-04-23
30	PRIOR APPLICATION NUMBER: 60/083366
31	PRIOR FILING DATE: 1998-04-27
32	PRIOR APPLICATION NUMBER: 60/083322
33	PRIOR FILING DATE: 1998-04-28
34	PRIOR APPLICATION NUMBER: 60/083392
35	PRIOR FILING DATE: 1998-04-29
36	PRIOR APPLICATION NUMBER: 60/083495
37	PRIOR FILING DATE: 1998-04-29
38	PRIOR APPLICATION NUMBER: 60/083466
39	PRIOR FILING DATE: 1998-04-29
40	PRIOR APPLICATION NUMBER: 60/083499
41	PRIOR FILING DATE: 1998-04-29
42	PRIOR APPLICATION NUMBER: 60/083545
43	PRIOR FILING DATE: 1998-04-29
44	PRIOR APPLICATION NUMBER: 60/083554
45	PRIOR FILING DATE: 1998-04-29
46	PRIOR APPLICATION NUMBER: 60/083558
47	PRIOR FILING DATE: 1998-04-29
48	PRIOR APPLICATION NUMBER: 60/083559
49	PRIOR FILING DATE: 1998-04-29
50	PRIOR APPLICATION NUMBER: 60/083500
51	PRIOR FILING DATE: 1998-04-29
52	PRIOR APPLICATION NUMBER: 60/083742
53	PRIOR FILING DATE: 1998-04-30
54	PRIOR APPLICATION NUMBER: 60/084366
55	PRIOR FILING DATE: 1998-05-05
56	PRIOR APPLICATION NUMBER: 60/084414
57	PRIOR FILING DATE: 1998-05-06
58	PRIOR APPLICATION NUMBER: 60/084411
59	PRIOR FILING DATE: 1998-05-06
60	PRIOR APPLICATION NUMBER: 60/084637
61	PRIOR FILING DATE: 1998-05-07
62	PRIOR APPLICATION NUMBER: 60/084639
63	PRIOR FILING DATE: 1998-05-07
64	PRIOR APPLICATION NUMBER: 60/084640
65	PRIOR FILING DATE: 1998-05-07
66	PRIOR APPLICATION NUMBER: 60/084598
67	PRIOR FILING DATE: 1998-05-07
68	PRIOR APPLICATION NUMBER: 60/084600
69	PRIOR FILING DATE: 1998-5-07
70	PRIOR APPLICATION NUMBER: 60/084627
71	PRIOR FILING DATE: 1998-05-07
72	PRIOR APPLICATION NUMBER: 60/084643
73	PRIOR FILING DATE: 1998-05-07
74	PRIOR APPLICATION NUMBER: 60/084643

```

; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Alignment Scores:
Pred. No.: 4,9e-34 Length: 3143
Score: 444.00 Matches: 153
Percent Similarity: 45.00% Conservative: 81
Best Local Similarity: 29.42% Mismatches: 200
Query Match: 11.81% Indels: 86
DB: Gaps: 31

US-09-874-198-2 (1-686) x US-09-978-697-168 (1-3143)
QY 194 G1G1G1uenseSerProgluTyrrProArgProTyrrProLysLeuSerSerCysThrTy 213
    ||| |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 1090 GGGGTCTCGACGACCCCGTACTTCCCGAGCTACTACGCCGCCAAACCCACTCTCTGG 1149
QY 214 SerIle-----SerLeugluGluGlyPheSerValIleLeuAspPheVal----- 228
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1150 CACCTGACGGTCCCTCTGAGTACGACGCTGCGCTGCTGTTGATGCGTATGACGTG 1209
QY 229 -----GluSerPheAspValIgluThrHisProgluThrLeuSerProTyrrAspPheLeu 246
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1210 AGGAGCGAGAAATGATGATTTGCGC-----TGCACCCAGGCCAGTGG 1251
QY 247 LysIleGlnThrAspArgGluGluHisGlyProPheCysGly-----LysThr 262
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1252 ACGATCCAGAACAGAG-----CTGTGTGCTTGGCGCATCTCGACGCC 1296
QY 263 LeuProHisArgIleGluThrLysSerAsnThr--ValThrIleThrPheValIThrAsp 281
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 297 TACGCCGAGAGATCCCGGTGGGCGCACGCGCGGATCCACATCACTTCACCTCCAG 1356
QY 282 GluSerGlyAspHisThrGlyTrpLysIleHisTy-----ThrSerThrAlaGlnPro 299
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1357 ATCTCCCTACCGCGCGCGGTGTGGGTCACATATGCTTGTACAAACAGTGGAGGCC 1416
QY 300 CysProTyrrPro---MetAlaProProAsnGlyHisValSerProValIleAlaLysTy 318
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1417 TGGCCCTGGAGAGTCTCTCTGTGTGATGAGACTGTGTCTCT-----GCCGTGAT 1470
QY 319 IleLeuLysAspSerPheSerIlePheCysGluThrGlyTrpGluLeuLeuGlnIleHis 338
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1471 GGGGTCAAGAGC-----TGCACCAAGCGC----- 1494
QY 339 LeuProLeuLysSerPheThrAlaValGlyGln-----LysAspGlySer 353
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1495 -----CTGGATGAGAGAAATCGGTTTGGAGGCGCACATTCAGTGCAGAAAGGAGCAG 1548
QY 354 TrpAspArgProMetProLysSerIleValAspCysGlyProProAspAlaPro 373
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1549 ACGTGCATCTCACTGCC-----AAGGTCTGTGAT---GGGACGCTTGATTTCTTC--- 1596
QY 374 SerGlyArgValGluTrpIleThrGlyProGlyValIThrThrTyrrLysAlaValIleGln 393

```

```

Db 1597 AACGCGACGATGAAGACAGTCCAGAGAGGGTCCCATGT-----GGACATTCACC 1650
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 394 TyrSerCysGluGluThrPheTyrrThrMetLysValAsn-----AspGlyLysTy 410
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1651 TTCCAGTGTGAGGACCGGAGCTCGGTGAAGAACCCACCGCATGTATGGCGGCC 1710
QY 411 ValCysGluAlaAspGlyPheTrpThrSerSerLysGlyGluLysSerLeuProValCys 430
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1711 GACTGCAG---GACGGC-----TCCGATGAGAGAGC-----TGT 1743
QY 431 GluProValCysGlyLeuSerAlaArgThrThrGlyValArgIleTyrrGlyGlnLys 450
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1744 GAC-----TGTGGCTTCAGAGGCCCTCCACG-----CGCATTTGTGTGAGCTGTG 1791
QY 451 AlaLysProGlyAspPheProTrpGln-----ValLeuIleLeuGlyGlyThrThrAla 468
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1792 TCCTCCGAGGGTGAAGGGCCATGGCAGGCCAGCTCCAGGTTGGGTCGACACATCTGT 1851
QY 469 AlaGlyAlaLeuLeuTyrrAspAsnTrpValIleuThrAlaAlaHisAlaValTyrrGln 488
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1852 GGGGGGGCCCTCATGCTGACCCCTGGGTGATACAGCTGCCACTGCTTCCAGAGAGAC 1911
QY 489 LysHisAspAlaSerAlaLeu--AspIleArgMetGlyThrLeuLysArgLeuSer--- 506
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1912 ACGATGGCTCCACGGGTGCTGTGACCGGTGTCCTGGCAGAGGTGTGAGAACCTCGCG 1971
QY 507 ---ProHisTyrrThrGlnAlaTrpSerGluAlaValPheIleHisGluGlyThrHis 525
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1972 TGGCTCGAGAGGTGCTTCAAGGTGAGCCGCGCTCTCTGACCGC---TACACAGAA 2028
QY 526 AspAlaGlyPheAspAsnAspIleAlaLeuIleLysLeuAsnLysValIleAsn 545
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2029 GAGGACAGCCATGACTACAGACGAGGGCTGTCTGACTGACACACCGGTGTGGCTCG 2088
QY 546 SerAsnIleThrProIleCysLeuProArgLysGluAlaGluSerPheMetArgThrAsp 565
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2089 GCCGCGTGGCGCCCTGTGCTGCTGCGCGCGCGC-----TCCACATTTCTGAGCGCGC 2142
QY 566 AspIleGlyThrAlaSerGlyTrpGly---LeuThrGlnArgGlyPheLeuAlaArgAsn 584
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2143 CTGCACTGCTGATTTACGGGTGGGCGCTTCGCGAGGGCGGCCCTTACAGACGCT 2202
QY 585 LeuMetTyrrValAspIleProIleValAspHisGlnLysCysThrAlaIleAlaTyrrGlyLys 604
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2203 CTGCAGAAAGTGAATGTGCAGTTGATCCACAGACCTGTGCAGCGACGCTAT----- 2256
QY 605 ProProTyrrProArgLysSerValIThrAlaAsnMetLeuCysAlaGlyLeuGlnSerGly 624
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2257 -----CGTACACAGGTGACGCGCACACGATGCTGTGTGCGGTACCGCAGAGGC 2304
QY 625 GlyLysAspSerCysArgGlyAspSerGlyAlaLeuValPheLeuAspSerGlyThr 644
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2305 AAGAAAGATGCCGTGACGAGGTGACTCAGGTGTGTCCTGGT---TGCAGAGACTCACT 2361
QY 645 GluArgTrpPheValGlyLysIleValSerTrpGlySerMetAsnCysGlyGluAlaGly 664
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2362 GGGCGCTGTGCTTGGCGGGGCTGTGCTGACGTGGGC---CTGGCGTGTGGCGGCTAAC 2418
QY 665 GlnTyrrGlyValTyrrLysValIleAsnTyrrIleProTrpIleGluAsnIleIleSer 684
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2419 TACTTCGGGCTGTACACCGCATCACAGGTGTGATGATGATGATCAGCAAGTGGTAC 2478

```

RESULT 24

```

US-09-978-192A-168
; Sequence 168, Application US/09978192A
; Patent No. US2002017753A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan

```

APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;
APPLICANT: Paonli, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C9
CURRENT APPLICATION NUMBER: US/09/978,192A
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/07450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/07632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30

PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559

Db 2257 -----CGTACCAGGTGACGCCACGCATGCTGTGCGCCGCTACCGCAAGGC 2304
Qy 625 GLYLysAspSerCysArgGlyAspSerGlyAlaLeuValPheLeuAspSerGluThr 644
Db 2305 AACAGAGATGGCTGCTAGGACTCAGGTGGTCCCTGGTG---TGCAAGACATCAGT 2361
Qy 645 GUAArgTrpPheValGlyGlyLeuValSerTrpGlySerMetAsnCysGlyGluAlaGly 664
Db 2362 GCGCGTGGTCTGCTGGCGGGGCTGCTGCTGCTGGG---CTGGGCTGTGGCCGCTAAC 2418
Qy 665 GUAArgGlyValThrTrpValIleAsnTrpIleProTrpIleGluAsnIleIleSer 664
Db 2419 TACTTGCGGCTACACCGCATCACAGTGTGATCATCAGCATGACGCAAGTGTGAC 2478
RESULT 25
us-09-999-832A-168
Sequence 168, Application US/0999832A
Serial INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kiljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James:
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acid Reference: P2630PlC63
CURRENT APPLICATION NUMBER: US/09/999,832A
PRIOR FILING DATE: 2001-10-24
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004

PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22

[illegible]

Pred. No.:	4..9e-34	Length:	3143
Score:	444.00	Matches:	153
Percent Similarity:	45.00%	Conservative:	81
Best Local Similarity:	29.42%	Mismatches:	200
Query Match:	11.81%	Indels:	86
DB:	9	Gaps:	31

US-09-874-198-2 (1-686) x US-09-999-832A-168 (1-3143)

QY	194	GLYGLULEUSERSERPROGLUTYRPROALGPROTYRPROLYSLEUSERSERCYSTHRTYR	213
DB	1090	GGCGTCCACACACCCTCTCCAGCTACCTACCTCCGCCCCAACACACCTACCTCTCGG	1149
QY	214	SERLE-----SERLEAGUGLUGLYPHESEVALILEUASPHEVAL-----	228
DB	1150	CACCTCAGGTGCCCTCTCTGGACTACGGCTTGCCCTTGCTTGATGCCATGCAGTGG	1209
QY	229	-----GLUSERPHEASPVALLGUTHRHISPROGLUTHRLEUCYSPROTYRASPHLEU	246
DB	1210	AGGAGGCGAAGATGATGATTGGCC-----TGACCCAGGCGCCATCGG	1251
QY	247	LYSLIEGLNTHRSAPRARGGLUGLHISGLYPROPHCYSGLY-----LYSTHR	262
DB	1252	ACGATCCGAACAGGAGG-----CTGTGGCTTGCCGCAATCCGAGGCC	1296
QY	263	LEUPROHISARGILEGLUTHRTHLYSSEASNTHR---VALTHRILETHRPHYVALTHRASP	281
DB	1297	TACGGCGAGAGGATCCCGGCGTGGCCAGCGCGGAGTACCATCATCATCCTACCTCCAG	1356
QY	282	GLUSERGLYASPHISHTHRTYRPLYSLIEHISTYR-----THSERTHALAEINPRO	299
DB	1357	ATCTCCCTCACCGGCGCGGTGTCGGGTGCATATGAGCTTGTAACACACAGTGGAGCCC	1416
QY	300	CYSPROTYRPRO---METALPROPROASNGLYHISVALSERPROVALGLINALYSTYR	318
DB	1417	TGCCCTGGAGAGTTCCTCTGTCTGTGAAAGACCTGTGTCCCT---GCCGTGAT	1470
QY	319	ILEULYASPSERPHESERILEPHECYSGLUTHRGITYRGLULEULEUNGILYHIS	338
DB	1471	GGGGCAAGGAC-----TGCCCCAAAGGCG-----	1494
QY	339	LEUPROLEULYSSERPHETHRILALVALCYGLN-----LYSAPGLYSER	353
DB	1495	-----CTGGATGAGAGAAACTGGCGTTGGAGAGCCACATTCAGTGCACAAAGAGACAC	1548
QY	354	TRPASPARGPROMETCPROALACYSSERILEVALASPCYSGLYPROPROASPHLEU	373
DB	1549	ACATGCATCTACATCGCC-----AAGCTCTGAT---GGGAGCGCTGATGTCTC--	1596
QY	374	SERGLYARGVALGLUTYRILETHRGYPROGLYVALTHRTYRILYSLAVALILEGLN	393
DB	1597	AACGCGACCGATGAGAAGCAGTGCAGGAAGGGGTCCATGT---GGGCAATTGCAC	1650
QY	394	TYRSERCYSGUGLUGUTHRPHETYRTHMETLYEVALASN-----ASPGLYSTYR	410
DB	1651	TTCCAGTGTGAGACCGGAGCTCGTGAAGAAGCCACACCGCAGCTGATGGGGGCC	1710
QY	411	VALCYSGULAAASPGLYPHETRPTHRSERSELYSGLYGLULYSSEULEUPROVALCYS	430
DB	1711	GACTGCACAG---GACGCG-----TCCGATGAGAGACAC---TGT	1743
QY	431	GLUPROVALCYSGLYLEUSERALARGTHRTHGLYARGILEYGLIYGLYGLNYS	450
DB	1744	GAC-----TGTGGCTCTCCAGGCGCCCTCCAGC---CGCATGTGTGTGTGACCTGTG	1791
QY	451	ALALYSPROGLYASPRPHEPROTYRGLN-----VALLEUILEUGLIGLYTHRTHALA	468
DB	1792	TCTCTCCAGGCTGAGTGCATGGCAGCGGCTCCAGGTTCCGGGTGCACACATCTGT	1851
QY	469	ALAEGLYALAEULEUTYRASPSANTPRVALLEUTHRALALAHISALAVALLTYRGLUGLN	488

Oy 489 LysHisAspIleAsnIleLeu---AspIleuGmeGlyThrLeuLysArgLeuSer--- 506
 Db 1912 AGCATGGCCCTCCACGGGTGCTGTGGACCCGTTCTGGGCAGAGTGTGCAGAACTCGGC 1971
 Oy 507 --ProHisTyrThrGlnAlaTrpSerGlnAlaValPheIleHisGluGlyTyrThrHis 525
 Db 1972 TGGCGCTGGAAGAGGAGTTCCTTCAGAGTGAGACCGCGCTGCTCTGCACCCG---TACAGAGA 2028
 Oy 526 AspAlaIleValPheAspAsnAspIleAlaLeuIleLysLeuAsnAsnLysValValIleAsn 545
 Db 2029 GAGAGACGCCATGACTACAGACGCGGGCGTGCAGACTCGACACACCCGCTGGTGGCGTCGC 2088
 Oy 546 SerAsnIleThrProIleCysLeuProArgLysGlnAlaGluSerPheMetArgThrAsp 565
 Db 2089 GCCCGCGTGGCCCCCGCTGCTGCTGCCCGCGCGC-----TCCCACTTTTGGACCGCGC 2142
 Oy 566 AspIleGlyThrAlaSerGlyTyrPgly--LeuThrGlnArgGlyPheLeuAlaArgAsn 584
 Db 2143 CTCGACATGCTGAGATTACGGGCTGGGGCGCTTGGCGAGAGGGCGGCCCATGACAGACCT 2202
 Oy 585 LeuMetLysValAspIleProIleValIleAspHisGlnLysCysThrAlaAlaTyrGluLys 604
 Db 2203 CTGCAGAAATGTGATGTGCAGTTGATCCCAAGACAGACTGTGCAGAGGAGGCTAT----- 2256
 Oy 605 ProProTyrProArgLysSerValThrAlaAsnMetLeuCysAlaGlyLeuGluSerGly 624
 Db 2257 -----GCTACACAGGTGAGCGCACCAACCATCTGTGGCGGTACCGCAGAGGC 2304
 Oy 625 GlyLysAspSerCysArgGlyAspSerGlyAlaLeuValPheLeuAspSerGluThr 644
 Db 2305 AAGAAGATGCTCTGTAGGGTGACTGAGGTGCTCCGCTGCTG---TGCAGGCACTCAGT 2361
 Oy 645 GluArgTrpPheValGlyLysIleValSerTyrPglySerMetAsnCysGlyGluAlaGly 664
 Db 2362 GGCCGCTGTGTTCTGTGGGGGGGCTGTACACTGGGGC---CTGGCGTGTGGCGGCGCTAAC 2418
 Oy 665 GlnTyrGlyValTyrThrLysValIleAsnTyrIleProTyrIleGluAsnIleIleSer 684
 Db 2419 TACTTCGGCGCTTACACCGCATACAGAGGTGTATGACTGGATCCACCAAGTGGTATGC 2478

THIS PAGE BLANK (USPTO)